Discovering the Dynamics of the Minimal Clinically Important Difference of Health Status Instruments in Patients with Chronic Obstructive Pulmonary Disease

Mol-Alma, Harma

DOI:
10.33612/diss.136223443

Discovering the Dynamics of the Minimal Clinically Important Difference of Health Status Instruments in Patients with Chronic Obstructive Pulmonary Disease

Harma Alma

Alma, Harma Johanna

Discovering the Dynamics of the Minimal Clinically Important Difference of Health Status Instruments in Patients with Chronic Obstructive Pulmonary Disease

PhD dissertation, University of Groningen, the Netherlands

ISBN print	978-94-034-2853-6
ISBN e-pub	978-94-034-2854-3
Lay-out	Guus Gijben, Proefschrift-AIO
Printing	Gilde Print
Cover design	Guus Gijben, Proefschrift-AIO
Cover image	Bayerisch Gmain, Berchtesgadener Land Tourismus GmbH

The cover image was selected for multiple reasons. First, it exemplified the environment afforded to patients undergoing pulmonary rehabilitation interventions for chronic obstructive pulmonary disease (COPD) in Bad Reichenhall, Germany, as discussed in this thesis. Second, the image reflected the most important quality of life aspects of the care setting mentioned by a patient with COPD, who was interviewed in Chapter 1. The patient stated that exposure to the mountains and clean air when walking around the lake were synonymous with a good health status. Third, the lake in the picture mirrors the mountains, contrasting the variation in highs and lows experienced by patients with COPD and their periods of better health and of frequent exacerbations. Finally, the mountainous regions of the Alps are a favoured site of the author of this thesis.

Printing of this thesis was financially supported by the University of Groningen, University Medical Center Groningen, Junior Scientific Masterclass, Graduate School of Medical Sciences (SHARE) and SBOH (employer of general practitioner trainees).

rijksuniversiteit
groningen

Discovering the Dynamics of the Minimal Clinically Important Difference of Health Status Instruments in Patients with Chronic Obstructive Pulmonary Disease

Proefschrift

ter verkrijging van de graad van doctor aan de
Rijksuniversiteit Groningen
op gezag van de
rector magnificus prof. dr. C. Wijmenga
en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op

woensdag 4 november 2020 om 12.45 uur

door

Harma Johanna Alma

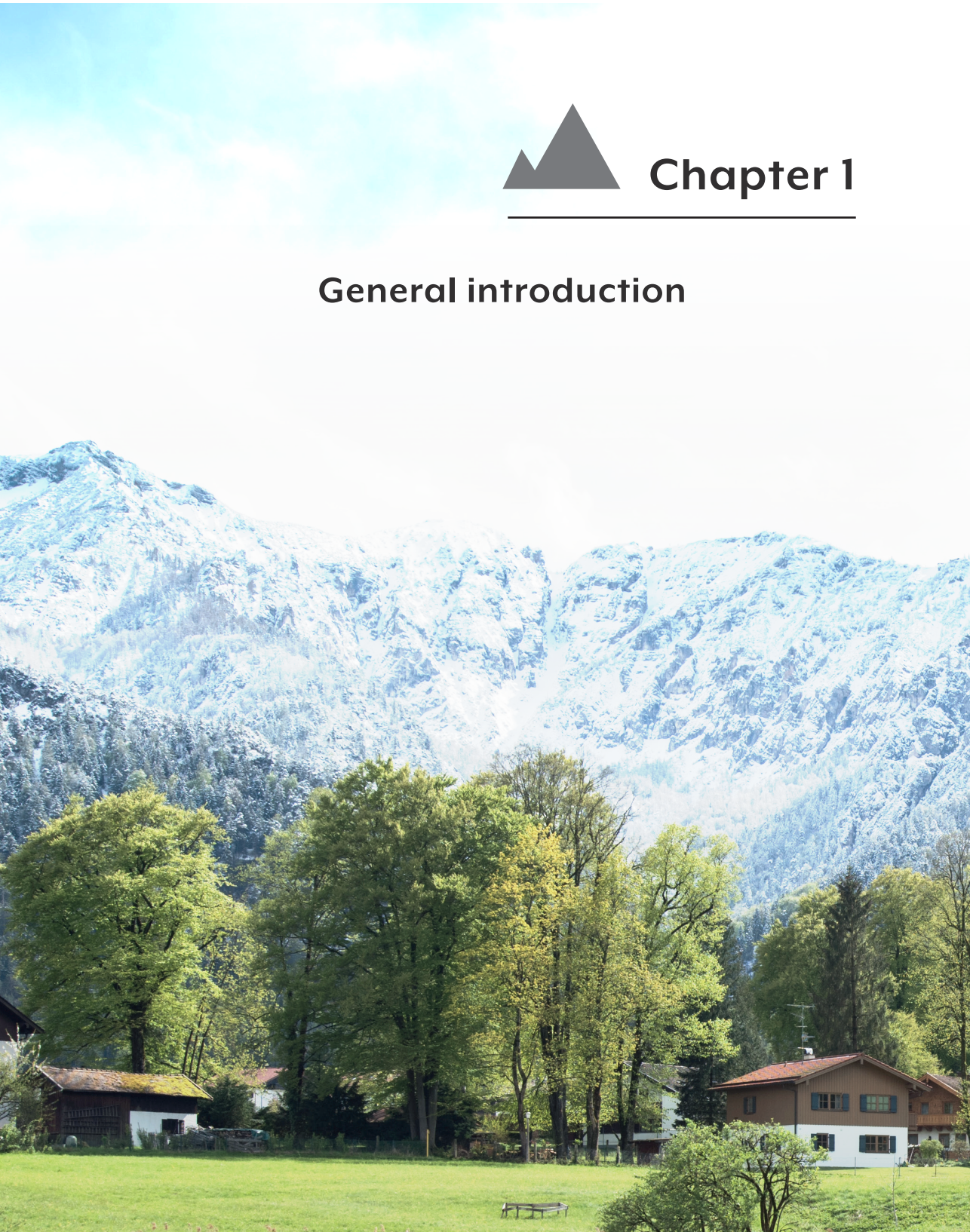
geboren op 31 maart 1984
te Coevorden

Promotores
Prof. dr. T. van der Molen
Prof. dr. R. Sanderman

Copromotor
Dr. C. de Jong

Beoordelingscommissie
Prof. dr. A.V. Ranchor
Prof. dr. H.A.M. Kerstjens
Prof. dr. D. Singh

Table of Contents



Chapter 1

General introduction

1.1 Background

Chronic or progressive diseases can significantly affect the lives, well-being and functional abilities of patients. This is exemplified in the interview of a patient with chronic obstructive pulmonary disease (COPD) attending pulmonary rehabilitation (PR) in Germany (*Box 1*). In this case, the chronic disease led to challenges with many daily activities and to multiple hospital admissions, both of which had severe and adverse effects. The impact of an illness on one's life must be measured, quantified, and interpreted, especially after an intervention. Although traditional physiologic parameters, such as spirometry, chest x-rays, oxygen saturations, and blood serum results provide important information to clinicians, these outcomes are often of less interest and importance to the patient. In addition, such parameters tend to have only a weak correlation with the patient's functional capabilities, experienced symptoms, and general well-being [1]. Indeed, patients with similar physiological outcomes can have major differences in their experienced quality of life (QoL), which has led to the integration of patient-reported outcomes (PROs) or questionnaires. Measuring QoL has become an obligatory outcome for the approval of new drug therapies during pharmaceutical trials [2-5], and is now used as a valid endpoint in randomised clinical trials evaluating PR and similar interventions. QoL measurement is also integrated in routine clinical practice (RCP) to validate new treatment regimens or to evaluate established guidelines [4]. However, an important challenge exists when *interpreting observed changes* in QoL among chronically ill patients enrolled in clinical practice, trials and interventions like PR (*Box 1*). The current thesis addresses the pivotal topic of measuring and interpreting changes in QoL during interventional research and routine medical care.

1.2 Measuring quality of life

1.2.1 Quality of life

QoL is a concept that is difficult to define and for which many definitions therefore exist, not least because it holds different meanings for different people [6-9]. As a general definition, one can state that "*QoL is the degree of satisfaction or dissatisfaction with various aspects of life that may be important to the individual*" [8]. However, it may also be considered "*the gap between what is desired by [an] individual and what is achievable [by an] individual*" [9]. QoL also includes the following broad range of considerations: functional capabilities and limitations in self-care, mobility, and physical activity; experienced (physical) symptoms and signs; the execution of role activities in one's work, personal life, and household management; the level of social functioning in personal interactions, intimacy, and communication; one's emotional status, including aspects of anxiety, stress, depression, control experienced and spiritual well-being; cognition status; the level of

Box 1: Interview with a patient with COPD during pulmonary rehabilitation in Bad Reichenhall, Germany

Interview 24th of July 2015: "In the Klinik Bad Reichenhall (Germany) I met a 63-year old female, who was admitted for an extended 3-week pulmonary rehabilitation programme. She was diagnosed with chronic obstructive pulmonary disease (COPD) in 2007. Her first symptoms were progressive coughing and dyspnoea that initially required a 3-week admission to the intensive care of the regional hospital. Her diagnosis was very severe COPD (grade IV) according to the global initiative for obstructive lung disease (GOLD). She also had severe osteoporosis and was using anti-depressive medication.

She reported smoking approximately two packs of cigarettes each day for 45 years and was aware that this had caused her COPD. She had quit smoking shortly after being diagnosed with the disease. Initially, her symptoms were stable and the disease was well managed, but since 2013 she had required frequent hospital admissions for exacerbations with extreme coughing, dyspnoea, and increased sputum production. Despite using multiple inhaled bronchodilators, she has since required daily oxygen 2 L/min by nasal cannulae. Initially treated with antibiotics and steroids, her frequent exacerbations were taken as justification for her referral to the pulmonary rehabilitation programme.

Before being diagnosed with COPD, she and her husband had owned a small bed and breakfast in the mountains near Oberstdorf in southern Germany. The pulmonary physician, who had diagnosed COPD, recommended that she sell her business and take care of herself, which she did. However, her life had changed dramatically since then. Her husband had died of cancer and she was currently living with her daughter, son-in-law, and two grandchildren. In her daily life, she experiences many COPD symptoms and functional limitations, such as household chores, self-care (e.g., washing, showering, getting dressed, and cooking), and driving a car. Her children help with many of these tasks. Mentally and physically she enjoys and benefits from short strolls around the local lake with the aid of her walker and oxygen, but needs to make frequent stops. She benefits significantly from the compassion and help of her family, friends, and neighbors, which gives her strength, especially during depressive episodes.

She is aware that COPD is progressive, and this contributes to her feeling down. Despite having tears in her eyes while talking about coping with COPD, I noticed an active will to fight for good quality of life and disease stability. She therefore enrolled in a randomised controlled clinical trial (RCT) investigating inspiratory muscle training (IMT) added to pulmonary rehabilitation. She explained that spirometry and a 6-minute walking test were performed at the start of her rehabilitation and that she had to fill out various questionnaires, but that she did not really understand why these were needed. Her assumption was that they were used for administrative purposes only.

The rehabilitation programme consisted of breathing exercises, physiotherapy, aerobics, walking training, GALILEO vibration board training, massage, IMT, and an educational programme (e.g., smoking cessation). An element that she enjoyed most were the daily walks around the local *Gradierwerk* in Bad Reichenhall, which she reported as a means of exercise that helped her breath easier. Her main goals were to lessen her dyspnoea, prevent recurrent exacerbations, and slow down disease progression. The ability to drive her car again was another major goal."

sleep and rest; one's experienced level of energy and vitality; general health perceptions; and one's overall life satisfaction [6-7, 10].

1.2.2 Health-related quality of life and health status

QoL is a much more comprehensive term than the concepts of either health-related quality of life (HRQoL) or health status, because it includes aspects of the individual's environment that may or may not be influenced by his or her health, illness or treatment [11-12]. By contrast, HRQoL mainly focuses on the "*health concepts or aspects of human life and activities that are generally affected by health conditions, illnesses, or health services*" [11]. It includes fundamental health dimensions, such as physical functioning, psychological and social functioning, performed role activities, overall life satisfaction, and perceptions of health [6, 13-14]. Health status is often considered to be the standardised outcome measure for HRQoL, and as such is frequently used with equivalence. Specifically, health status has been defined as *"the impact of health (or disease) on a person's ability to perform and derive fulfilment from the activities of daily life"* [11-13]. The concept of functional status is also frequently considered in the context of health status and HRQoL, and is defined as "*a person's ability to perform a variety of physical, emotional, and social activities"* [11-13]. However, using the concepts of QoL, HRQoL, health status, and functional status interchangeably can lead to confusion with the terminology [6, 10].

1.2.3 The measurement of health status

Health status measurement is a standardised way of quantifying and scoring the impact of health or disease on a patient's life, health, and well-being [2]. Various tools have emerged over recent decades [1, 4-5, 9-11, 13, 15]. Disease-specific measures include elements of health status relevant to a given population [9, 15]. These tools are more likely to be shorter, and the measures are sensitive for the specific health problems of the disease. Alternatively, general health status assessments can be used to compare HRQoL between patient populations and tend to be robust thanks to long development and testing phases [15]. Health status measures can range from a single question to a complex combination of questions [1, 5-6], and can include a single score or multiple subscores for specific health status domains (e.g., symptoms, functional status, and emotional well-being) [1, 5-6, 10, 15]. The measures can be assessed in one of several ways [1].

Measures administered directly by the patient are defined as the so-called PROs [11, 16]. These questionnaires capture an individual's experience of the impact of health or disease, without the interpretation of others [5, 16, 17-19]. These instruments are of major interest, because some treatment effects are known only to the patient and not to the physician or researcher [5]. PROs provide a unique perspective on treatment effectiveness. Formal and standardised evaluation of QoL by PROs may be more reliable than informal patient

interviews. PROs usually measure concepts of overall health status, functional status (including daily activities and exercise capabilities), disease symptoms and signs, health perceptions, treatment satisfaction, and treatment adherence [5, 11]. At a minimum, they should include components of physical, psychological and social functioning [5].

1.3 Defining the importance of change in health status

1.3.1 Concepts of change

An important consideration of *any assessment* is how the physician or researcher should score and interpret whether important change occurred in a patient's disease, QoL or experienced health status after a therapeutic intervention (e.g., pulmonary rehabilitation, *Box 1*). A given patient will tend to have specific goals that they wish to achieve when engaging in an intervention, and these should be captured, scored and evaluated in a standardised manner. Along with traditional physiological parameters, health status is now routinely captured as an obligatory endpoint of treatment. It is for this reason, not for mere administrative purposes, that patients are required to complete the various PROs during research and therapy.

A health status questionnaire or instrument should have accurate psychometric properties to be deemed applicable for use in evaluating interventions. These important characteristics include its appropriateness, reliability, validity, practicality, interpretability and responsiveness [1, 5-6, 11, 15, 20]. Responsiveness has been defined as *"the ability of an instrument to detect change when it occurs"* [21-22]. Random changes (i.e., intervening *noise*) in scores will always occur from before to after an intervention, but do not necessarily indicate real or important change for the patient (i.e., a true *signal*) [1].

The concept of change can be considered in different ways, and these have been grouped into specific levels or categories [21, 23]. First, change can be considered as either occurring *within* the same individual, or as a difference *between* individuals [21, 23]. Second, change can be considered at a group level with multiple patients, or at an individual level with a single patient. Third, change can be either positive (indicating *improvement*) or negative (indicating *deterioration*).

The definition of change also differs by health status instrument and should be considered in clinical practice [21]. The first level is the *minimum potentially detectable change by an instrument*, which is the amount of change that an instrument can measure at the minimum level without interpreting the reality and importance of the measured change score. The second level is the *minimum change detectable given the measurement error of*

the instrument, which includes the minimum amount of change that an instrument can measure beyond any random variation in scoring (i.e., measurement error or *noise*). The third level is the *observed change measured by the instrument in a given population*, which describes the real observed change (i.e., *true signal*) measured in a group. Last, the fourth level is the *observed change in a population deemed to have improved by either the patient, clinician or other*. It includes the real observed change (i.e., *true signal*) measured in a group, but also incorporates a value statement of this change either by the patient, clinician, or someone else.

1.3.2 The concept of minimal clinically important difference

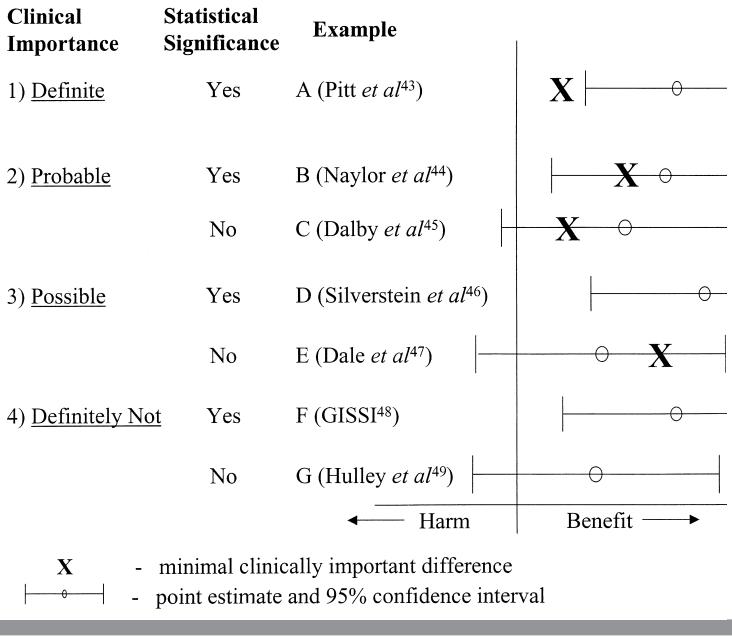
Change in health status should be proven to be statistically significant in clinical trials, based on a low chance of the observed difference resulting from pure chance (usually at the 5% threshold). It is well known that the significance of such results is affected by the sample size [23-24], with larger cohorts being more likely to identify even small changes or differences as statistically significant. However, this does not necessarily mean that the results are also clinically important or relevant to the patient [21, 23-27]. This led to Jaeschke *et al*. developing the minimal clinically important difference (MCID) in 1989, defined as *"the smallest difference in score in the domain of interest, which patients perceive as beneficial and which would mandate, in the absence of troublesome side effects and excessive cost, a change in the patient's management"* [28].

Many alternative wordings have emerged since this original definition for the same intended concept as the MCID, but these are not always synonymous. Terms include, among others, the minimum important difference (MID), subjectively significant difference (SSD), clinically important difference (CID), minimal clinical difference (MCD), minimally detectable difference (MDD), minimally important change (MIC), minimum detectable change (MDC), minimal clinically significant difference (MCSD), minimally perceptible difference (MPD) and many more [29-31]. It could be argued that any *difference* refers to change *between* groups while *change* itself can refer to change *within* a group. The persistence of interchangeable use of this terminology in the literature makes the matter less clear.

The MCID is a very important parameter for interpreting health status measurement in clinical trials and research. It is used to determine the required sample size in a scientific study based on health status as the primary outcome [25]. In addition, the MCID is used to *interpret* health status changes as an obligatory endpoint in many clinical trials, while clinicians may also use the MCID to guide care and set national guidelines. Most clinical trials in the European Union (EU) now require that a health status tool be included as a primary or secondary outcome parameter [2].

The MCID of an instrument is used to measure the threshold for clinically relevant change when interpreting results, and it is required that over 50% of patients in the trial group meet the MCID threshold for the intervention to be considered clinically important [17]. Alternatively, the mean change from pre- to post-intervention should be larger than the MCID of the tool. A schematic presentation of how to interpret treatment outcomes based on the MCID of an instrument is shown in *Figure 1* [32]. This shows the mean group result and its 95% Confidence Interval (95%CI), together with the significance of the trial outcomes, as evaluated with respect to the MCID of a tool. The goal is to determine the clinical importance of the observed change in the trial.

Figure 1: Interpreting the importance of clinical trial results with the MCID as a pivotal parameter (Published in Man-Son Hing, 2002 [32], permission for printing requested)



1.3.3 Methods of determining an instrument's MCID

There is currently no standard method to determine health status instruments' MCIDs [5, 30], but they tend to be calculated either as absolute or relative (percentage) estimates [23-24], even though there may be no real difference between these outcomes [33]. Overall, the methods can be divided into *anchor-*, *distribution-* and *opinion-based methods* (*Table 1*) [5, 13, 22-24, 26, 31, 34-45]. The recommended approach is to use several anchor-based methods with relevant clinical or patient indicators as the anchors, and examining various distribution-based estimates as supportive information [5, 41-42, 46]. Irrespective

of the method used, the patient perspective on the importance of the change should be given greatest weighting. Triangulation on a single value or a range can be performed by systematic review and/or Delphi procedures to finalise the MCID process, which should involve synthesizing clinical, statistical and qualitative data [42, 46].

Table 1: Overview of generally used methodology in determining a health status instrument's MCID

Anchor-based methods	Distribution-based methods	Opinion-based methods
Using a **global rating of change (GRC) scale** as an anchor (either patient or clinician rated): - Mean change score of the minimally changed population - Receiver operating characteristics (ROC) curves to determine the cutoff point for the minimally changed population - Regression analysis between health status instrument and GRC anchor	Using a **rule of thumb** (6%-10% of the maximum score of the instrument's scale)	**Delphi rounds of discussion** by experts in the field of interest or estimates based on expert opinions
Using a **correlated health status questionnaire or other clinical instrument** with a known MCID as the anchor: - Mean change score of the minimally changed population - Receiver operating characteristics (ROC) curves to determine the cutoff point for the minimally changed population - Regression analysis between health status instrument and anchor	Determining the **half standard deviation (0.5SD)** / **effect size (ES)** as equivalent to the MCID	
Selecting **a clinical event or comparing disease severity states between patients** as an estimate for the MCID	Determining the **standardised error of measurement (SEM)** as equivalent to the MCID	
Using **preference-based ratings between individual patients** as equivalent of the MCID	Determining the **reliability of change index (RCI)** as equivalent to the MCID (≥1.96)	

Abbreviations: ES, effect size; GRC, global rating of change; MCID, minimal clinically important difference; RCI, reliability of change index; ROC, receiver operating characteristics; SD, standard deviation; SEM, standard error of measurement.

Anchor-based methods use an external criterion (*the anchor*) as the reference point when quantifying changes or differences measured on a health status instrument [5, 13, 22-24, 26, 29, 31, 34, 36-44]. Many different anchors could be used as such an external criterion. These may include a patient's global rating of change (GRC) scale – also called transition rating (TR) scale, global perceived effect (GPE) scale, global impression of change (GIC) scale – to assess both the *within-person* change and the *between-person* change [5, 13, 22-25, 28-29, 31, 34-37, 40-48]. In this way, the MCID can be considered the mean change score among patients indicating small (minimal), yet important, change on the anchor instrument [5, 13, 22-25, 28-29, 34-37, 40, 45, 47]. However, one may also define the MCID as the mean difference

in score between patients indicating no change and patients indicating a small change on an anchor question [20, 22, 49-51]. Other anchors that can function as an external criterion include correlated clinical instruments with a known MCID, such as other established health status questionnaires [5, 13, 24, 29, 34-35, 42]. Statistical techniques like receiver operating characteristics (ROC) curves and/or regression analysis may also help to further clarify the MCID threshold (*Table 1*) [23-24, 43].

Another anchor-based method is to compare groups by disease severity and their health status scores [13], which represent the analysis of *between-patient* differences. An external criterion (e.g., numbers of doctors' appointments, exacerbations, or hospital admissions) may serve as anchors for differentiating clinical relevance between groups [13, 35]. Patients may also compare themselves (preference ratings) to others for clinically relevant differences in health status [13, 22, 34-35, 44-45, 49-51], or clinicians may be asked to complete prognostic ratings for their patients to evaluate therapy effects [13]. Regardless of the anchor-based method used, it is of major importance that there is a good correlation between the health status tool and the selected anchor, with correlations (*r*) preferably exceeding 0.30 (or even 0.50) [41-42]. Anchors should also be selected based on their relevance to the disease, clinical acceptance, validity, and existing evidence [42].

Distribution-based methods use different statistical parameters to assess clinical significance [5, 13, 23-24, 26, 31, 36-37, 40-45, 52]. These include the use of *Cohen's effect sizes (ES)*, *standardised response means (SRM)* and *standardised mean differences (SMD)*, with 0.30–0.50 standard deviations (SD) considered equivalent to the MCID [13, 22, 24, 40-44, 52]. Cohen's ES used to determine MCIDs vary from 0.20 for a small change, to 0.50 for a moderate change, and to 0.80 for a large change [13, 22-23, 29, 38, 52-53]. An estimate of 0.50SD turned out to reflect an instrument's MCID consistent with the results of anchor-based methods [5, 13, 22-23, 26, 31, 34, 38, 41-43]. As a *rule of thumb* (empirical rule), a 6%–10% change in the total score of an instrument is considered to approximate the MCID (*Table 1)* [5, 24].

The standard error of measurement (SEM) is *an alternative distribution-based method* that is used to determine the MCID. The SEM describes the error (i.e., *noise)* associated with an instrument [13, 25], representing the variation in the scores caused by the unreliability of the scale or measure [13, 22-23, 34, 40, 43-44, 54-55]. It is estimated as the SD of the instrument multiplied by the square root of one minus its reliability coefficient [54-55]. Although a SEM of one to two should be equivalent to an anchor-based MCID [13, 22-23, 44-45, 54-55], the one-SEM criterion has been shown to equate to a 0.50-point change on a 7-point scale [22, 38]. Next, the reliability of change index (RCI) can be calculated as the individual's change score divided by the square root of the SEM [13, 23, 56). If the RCI is larger than 1.96,

the change can be considered a true change within the 95% confidence level. Other statistical parameters include the smallest detectable difference (SDD) (i.e., the minimum detectable difference / change (MDD/MDC)), which is the smallest difference that an instrument can detect beyond the random error of measurement [23, 43, 57]. It may not necessarily reflect the MCID of an instrument, as it carries no impact of importance [43], but the MCID should at least exceed this level if the result is to make sense.

Opinion-based approaches are the third and final category of methods for determining the MCID [25, 36-37, 58]. In such an approach, experts are asked to evaluate and determine what would constitute clinically relevant changes based on their experience and case studies, using the Delphi method over several rounds of discussion [25, 31, 45, 58]. This is most often the stage used to finalise MCID determinations.

1.3.4 Advantages and disadvantages of the different methods

Anchor-based methods have a clear link with clinical practice and often involve the patient's own judgement of the importance and relevance of the experienced change after intervention [58]. The specific use of GRCs as anchors has strengths and weaknesses [48]. On the one hand, they are simple, easy to administer and interpret, demonstrate good validity, and involve the patient's perspective. On the other hand, the approach is heavily reliant on patient recall of their current and previous health (i.e., retrospective or interpersonal ratings) [51]. GRCs may thus be influenced by this so-called recall bias and, as such, may correlate more with the current health state than the former state. Furthermore, GRCs may not be reproducible and might be too global in nature [48].

Another important issue when using general anchor-based methods is that there may not be a linear relationship between the health status instrument at stake and the selected anchor [13, 34]. Using different anchors may also lead to different MCID estimates [13], possibly creating a range of values rather than a single convenient estimate for use in clinical practice [23, 36-37]. Moreover, MCIDs should always exceed the measurement error of the instrument, and this is not considered with anchor-based approaches only [23, 25].

An advantage of the distribution-based methods is that they provide a quick and easy way to establish change beyond a defined level of random variation [13]. However, a notable disadvantage is that there are few agreed-upon benchmarks for establishing clinically significant improvement [13]. Statistical measures of variation may also be larger in a more heterogeneous sample and may produce therefore a higher MCID [34]. Finally, these methods do not provide a sense of the clinical relevance of either the change or of patient involvement [13, 40, 43, 46].

1.4 Problems in determining the MCID

Assessing the MCID requires that one identifies the smallest changes that are important to patients, their families and their clinicians [11]. Repeated use of health status instruments and their MCIDs should lead to evolution over time. Given that health status is central to the assessment and management of chronic diseases like COPD, it is therefore crucial that their MCIDs are investigated thoroughly. Currently, however, the evidence for the MCIDs of most health status tools is limited, despite their continued use in scientific research and clinical practice. Several issues regarding the MCIDs of health status questionnaires have not been extensively investigated, which risks over- or underestimation of therapeutic effects. As such, interpretation of treatment outcomes in scientific trials and clinical practice based on these thresholds should be made with caution.

1.4.1 Lack of a standard approach for determining the MCID

Many definitions and methods are available that rely on patients, physicians and/or statistical analysis to evaluate clinically relevant changes in health status scores. However, the extent to which the various methods, anchors and statistical manipulations affect the MCID estimate for a health status instrument are yet to be clarified [13]. Researchers are currently free to (mis)use the diverse range of anchors and techniques to define an instrument's MCID. One would require comparisons of the available methods [22-23, 25-26, 30, 34], most likely resulting in the identification of a range of MCIDs for use in clinical practice [23, 36-37, 42].

1.4.2 Impact of follow-up period in measuring change and the MCID

The follow-up (recall) period, during which change is measured, may also affect an instrument's MCID [13, 24, 30, 34, 42, 48, 59-60]. GRCs assessed by patients or clinicians may be more closely related to current follow-up scores than to the original baseline health states. The longer the follow-up period, the harder it may be for patients and clinicians to recall a previous health state, causing possible recall bias. Patients correlating their rating of the importance of a perceived change with their current health state may be considered a response shift [49].

1.4.3 Direction of change and the MCID

Another issue is that MCIDs are mostly determined in settings where the patient has received a specific intervention (e.g., PR or pharmacotherapy) to improve his or her health status. It remains unclear whether MCIDs differ between cases of improvement and deterioration [13, 25-26, 34]. Chronically ill patients deteriorate over time, especially in cases of COPD [4, 61-62], so it may be that halting further disease progression is an important clinical outcome too. Thus, MCIDs for deterioration may be of clinical importance, but are yet to be studied.

1.4.4 Impact of disease severity state and context on the MCID

The MCID may depend on various factors that underly both disease severity and the health status dimension being measured [13, 23-25, 30, 34, 63-64]. General and disease-specific instruments have shown worse health statuses among patients with more severe disease stages [65], and it is unknown if one requires multiple MCIDs based on disease severity [17]. Worse health status scores may trigger exacerbations and hospital admissions, and as such, possibly require smaller changes to be clinically relevant; however, there will also be more room for improvement among these patients. This could potentially influence the relevant MCID of an instrument. A related area of concern is that an instrument's MCID may also be affected by the baseline health status score [13, 17, 23-25, 30, 34, 63-64]. This has also not been investigated in (COPD) MCID health status research.

Another unexplored concern is that MCIDs may be context-specific, differing by the study population and setting [5, 24-25, 27, 29, 41-42, 66]. MCIDs for interventions like PR or pharmacotherapy might differ when compared to routine medical care, or to patients during and/or following an exacerbation. Moreover, certain patient characteristics could influence the MCID. For example, age is known to affect health status [67], with younger patients being more impaired by chronic diseases like COPD [62] and tending to report worse health status scores, not least because symptoms have a greater impact on their function [68]. Female sex has also been associated with more exacerbations [69-70], possibly resulting in greater impairment of health status, and thus, potentially require different MCIDs.

1.4.5 Group MCIDs and interpretation of individual change

A challenging issue remains that MCIDs are mostly determined at the group level where significant variation exists between individual patients. Regression to the mean occurs, because MCIDs represent a group estimate in which extreme change scores are balanced by a greater number of average change observations [13, 30]. Consequently, it may be difficult to make a judgement about an individual's change in health status, which is a major area of interest for physicians in clinical practice [26-27, 66].

1.5 Research objectives and thesis outline

The problems that exist in measuring and applying MCIDs for health status instruments (*Paragraph 1.4*) form the foundation of the research questions posed in this thesis. There is a focus on health status tools for COPD, because this chronic disease is a leading cause of morbidity and mortality worldwide, and the concept of health status is well integrated in its assessment.

1.5.1 Research objectives

The main research goals of this thesis are *to analyse the general dynamics of determining the MCID for health status tools in a COPD context, and to develop an integrated system for use of these estimates in clinical practice and scientific research by making use of multiple methods over various time periods and settings, while considering factors of importance in doing so*. Based upon these aims and the defined problems with MCIDs, the following objectives have been set for this thesis:

- To examine and judge, in a systematic manner, the current evidence for the MCIDs of health status tools used for COPD (*Chapter 3*);
- To investigate the impact of selecting different methods, statistics, and anchors on the resulting MCIDs of health status tools for COPD (*Chapter 4*);
- To explore the impact of the recall period and measurement period on the MCID for health status tools for COPD (*Chapter 5*);
- To compare the MCIDs of health status tools for COPD when assessing improvement versus deterioration scores (*Chapter 6*);
- To establish an idea of the importance of patient-related factors in setting the MCID for health status tools for COPD (*Chapter 7*);
- To quantify the effect of baseline health status and disease severity on the MCID for health status tools for COPD (*Chapter 7*); and
- To provide an integrated framework for determining a health status instrument's MCID and its application in scientific research and clinical practice (*Chapter 8*).

Data for this thesis derive from two main studies. Study one comprises data from a randomised controlled clinical trial (RCT) on inspiratory muscle training (IMT) added to a 3-week PR programme for patients with COPD at the Klinik Bad Reichenhall, Germany [71] (see *Box I* for a representative case). Study two comprises data from routine clinical practice (RCP) for patients with COPD managed in primary and secondary care in the Netherlands. In this second study, patients received no specific intervention beyond standard care, per the Dutch COPD treatment guidelines. The primary goal was to measure health status changes during a 12-month period.

1.5.2 Thesis outline

This thesis has the following outline. *Chapter 2* presents additional background information on COPD and the integration of health status assessment in its management. *Chapter 3* summarises the procedures and results of a systematic review on health status instruments used in patients with COPD and the existing evidence for their MCIDs, providing quantitative and qualitative analyses with a final data synthesis. *Chapter 4* focuses on determining the MCID for the recommended health status tools for COPD using a variety of techniques, statistics and anchors. In *Chapter 5*, the results of the investigation into the extent to which the MCIDs of health status tools for COPD change when measured over different periods and with different anchor transition rating scales (e.g., GRCs) are presented. *Chapter 6* then details the results of whether MCIDs differ for improvement and deterioration based on data from PR and RCP. In *Chapter 7*, there is a discussion of the effects of various external factors on the measurement of MCIDs for health status tools for COPD. This includes the effects of the baseline health status score, disease severity, study context and patient-related characteristics on the MCID estimate for both improvement and deterioration. *Chapter 8* then presents a summary of the main results of this thesis and continues with a discussion of the overall findings. The goal is to synthesise the findings to produce guidelines for creating future MCIDs and applying them to health status tools in clinical practice when managing COPD.

1.6 References

[1] Guyatt, GH, Feeny, DH and Patrick, DL. Measuring health-related quality of life. *Annals of Internal Medicine* 1993; 118(8): 622-629.

[2] Jones, PW. Health status measurement in chronic obstructive pulmonary disease. *Thorax* 2001; 56(11): 880-887.

[3] Jones, PW. Health status and the spiral of decline. *COPD: Journal of Chronic Obstructive Pulmonary Disease* 2009; 6(1): 59-63.

[4] Jones, PW. Health Status: What Does It Mean for Payers and Patients? *Proceedings of the American Thoracic Society* 2006; 3(3): 222-226.

[5] US Department of Health and Human Services FDA Center for Drug Evaluation and Research. Guidance for industry: patient-reported outcome measures: use in medical product development to support labelling claims: draft guidance. *Health and Quality of Life Outcomes* 2006; 4: 79.

[6] Fitzpatrick, R, Davey, C, Buxton, MJ *et al*. Evaluating patient-based outcome measures for use in clinical trials. *Health Technology Assessment* 1998; 2(14): 1-74.

[7] Bullinger, M. Assessing health related quality of life in medicine. An overview over concepts, methods and applications in international research. *Restorative Neurology and Neurosciences* 2002; 20 (3-4): 93-101.

[8] Farquhar, M. Definitions of quality of life: a taxonomy. *Journal of Advanced Nursing* 1995; 22(3): 502-508.

[9] Jones, PW. Issues Concerning Health-Related Quality of Life in COPD. *Chest* 1995; 107(5 Suppl): 187S-193S.

[10] Bergner, M. Quality of Life, Health Status and Clinical Research. *Medical Care* 1989; 27(3 Suppl): s148-156.

[11] Curtis, JR and Patrick, DL. The assessment of health status among patients with COPD. *European Respiratory Journal Supplement* 2003; 41: 36s-45s.

[12] Reardon, JZ, Lareau, SC and ZuWallack, R. Functional Status and Quality of Life in Chronic Obstructive Pulmonary Disease. *The American Journal of Medicine* 2006; 119(10 Suppl 1): S32-37.

[13] Crosby, RD, Kolotkin, RL and Williams, GR. Defining clinically meaningful change in health-related quality of life. *Journal of Clinical Epidemiology* 2003; 56(5): 395-407.

[14] Patrick, DL and Bergner, M. Measurement of Health Status in the 1990s. *Annual Review of Public Health* 1990; 11: 165-183.

[15] Bergner, M and Rothman, ML. Health Status Measures: An Overview and Guide for Selection. *Annual Review of Public Health* 1987; 8: 191-210.

[16] Jones, P, Miravittles, M, van der Molen, T *et al*. Beyond FEV1 in COPD: a review of patient-reported outcomes and their measurement. *International Journal of COPD* 2012; 7: 697-709.

[17] Jones, PW, Rennard, S, Tabberer, M *et al*. Interpreting patient-reported outcomes from clinical trials in COPD: a discussion. *International Journal of COPD* 2016; 11: 3069-3078.

[18] Cazzola, M, Hanania, NA, MacNee, W *et al*. A review of the most common patient-reported outcomes in COPD – revisiting current knowledge and estimating future challenges. *International Journal of COPD* 2015; 10: 725-738.

[19] Haughney, J and Gruffydd-Jones, K. Patient-centred outcomes in primary care management of COPD - what do recent clinical trial data tell us? *Primary Care Respiratory Journal* 2004; 13(4): 185-197.

[20] Davidson, M and Keating, J. Patient-reported outcome measures (PROMS): how should I interpret reports of measurement properties? A practical guide for clinicians and researchers who are not biostatisticians. *British Journal of Sports Medicine* 2014; 48(9): 792-796.

[21] Beaton, DE, Bombardier, C, Katz, JN and Wright, JG. A taxonomy for responsiveness. *Journal of Clinical Epidemiology*, 2001; 54(12): 1204-1217.

[22] Hays, RD, Farivar, SS and Liu, H. Approaches and Recommendations for Estimating Minimally Important Differences for Health-Related Quality of Life Measures. *COPD: Journal of Chronic Obstructive Pulmonary Disease* 2005; 2(1): 63-67.

[23] Copay, AG, Subach, BR, Glassman, SD *et al*. Understanding the minimum clinically important difference: a review of concepts and methods. *The Spine Journal* 2007; 7(5): 541-546.

[24] Angst, F, Aeschlimann, A and Angst, J. The minimal clinically important difference raised the significance of outcome effects above the statistical level, with methodological implications for future studies. *Journal of Clinical Epidemiology* 2017; 82: 128-136.

[25] Beaton, DE, Boers, M and Wells, GA. Many faces of the minimal clinically important difference (MCID): a literature review and directions for future research. *Current Opinion in Rheumatology* 2002; 14(2): 109-114.

[26] Hays, RD and Woolley, JM. The Concept of Clinically Meaningful Difference in Health-Related Quality of Life Research. How Meaningful is it? *Pharmacoeconomics* 2000; 18(5): 419-423.

[27] Troosters, T. How important is a minimal difference? *European Respiratory Journal* 2011; 37(4): 755-756.

[28] Jaeschke, R, Singer, J and Guyatt, GH. Measurement of health status. Ascertaining the Minimal Clinically Important Difference. *Controlled Clinical Trials* 1989; 10(4): 407-415.

[29] Chapman, KR, Bergeron, C, Bhutani, M *et al*. Do We Know the Minimal Clinically Important Difference (MCID) for COPD Exacerbations? *COPD: Journal of Chronic Obstructive Pulmonary Disease* 2013; 10(2): 243-249.

[30] Cook, CE. Clinimetrics Corner: The Minimal Clinically Important Change Score (MCID): A Necessary Pretense. *Journal of Manual & Manipulative Therapy* 2008; 16(4): 82-83.

[31] Sloan, JA. Assessing the Minimally Clinically Significant Difference: Scientific Considerations, Challenges and Solutions. *COPD: Journal of Chronic Obstructive Pulmonary Disease* 2005; 2(1): 57-62.

[32] Man-Son Hing, M, Laupacis, A, O'Rourke, K *et al*. Determination of the Clinical Importance of Study Results: A Review. *Journal of General Internal Medicine* 2002; 17(6): 469-476.

[33] Zhang, Y, Zhang, S, Thabane, L *et al*. Although not consistently superior, the absolute approach to framing the minimally important difference has advantages over the relative approach. *Journal of Clinical Epidemiology* 2015; 68(8): 888-894.

[34] Dworkin, RH, Turk, DC, Wyrwich, KW *et al*. Consensus Statement Interpreting the Clinical Importance of Treatment Outcomes in Chronic Pain Clinical Trials: IMMPACT Recommendations. *The Journal of Pain* 2008; 9(2): 105-121.

[35] Jones, PW. Interpreting thresholds for a clinically significant change in health status in asthma and COPD. *European Respiratory Journal* 2002; 19(3): 398-404.

[36] Make, B. How Can We Assess Outcomes of Clinical Trials: The MCID Approach. *COPD: Journal of Chronic Obstructive Pulmonary Disease* 2007; 4(3): 191-194.

[37] Make, B, Casaburi, R and Leidy, NK. Interpreting Results from Clinical Trials: Understanding Minimal Clinically Important Differences in COPD Outcomes. *COPD: Journal of Chronic Obstructive Pulmonary Disease* 2005; 2(1): 1-5.

[38] Norman, GR. The Relation Between the Minimally Important Difference and Patient Benefit. *COPD: Journal of Chronic Obstructive Pulmonary Disease* 2005; 2(1): 69-73.

[39] Norman, GR, Sloan, JA and Wyrwich, KW. Interpretation of Changes in Health-related Quality of Life. The Remarkable Universality of Half a Standard Deviation. *Medical Care* 2003; 41(5): 582-592.

[40] Ostelo, RWJG, Deyo, RA, Stratford, P *et al*. Interpreting Change Scores for Pain and Functional Status in Low Back Pain. Towards International Consensus Regarding Minimal Important Change. *Spine* 2008; 33(1): 90-94.

[41] Revicki, DA, Cella, D, Hays, RD *et al*. Responsiveness and minimal important differences for patient reported outcomes. *Health and Quality of Life Outcomes* 2006; 4: 70.

[42] Revicki, D, Hays, RD, Cella, D *et al*. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *Journal of Clinical Epidemiology* 2008; 61(2): 102-109.

[43] Turner, D, Schünemann, HJ, Griffith, LE *et al*. The minimal detectable change cannot reliably replace the minimal important difference. *Journal of Clinical Epidemiology* 2010; 63(1): 28-36.

[44] Wyrwich, KW and Wolinsky, FD. Identifying meaningful intra-individual change standards for health-related quality of life measures. *Journal of Evaluation in Clinical Practice* 2000; 6(1): 39-49.

[45] Wells, G, Beaton, D, Shea, B *et al*. Minimal Clinically Important Differences: Review of Methods. *The Journal of Rheumatology* 2001; 28(2): 406-412.

[46] Leidy, NK and Wyrwich, KW. Bridging the Gap: Using Triangulation Methodology to Estimate Minimal Clinically Important Differences (MCIDs). *COPD: Journal of Chronic Obstructive Pulmonary Disease* 2005; 2(1): 157-165.

[47] Juniper, EF, Guyatt, GH, Willan, A *et al*. Determining a Minimal Important Change in a Disease-Specific Quality of Life Questionnaire. *Journal of Clinical Epidemiology* 1994; 47(11): 81-87.

[48] Kamper, SJ, Maher, C and Mackay, G. Global Rating of Change Scales: A Review of Strengths and Weaknesses and Considerations for Design. *Journal of Manual & Manipulative Therapy* 2009; 17(3): 163-170.

[49] Redelmeier, DA, Guyatt, GH and Goldstein, RS. Assessing the Minimal Important Difference in Symptoms: A Comparison of Two Techniques. *Journal of Clinical Epidemiology* 1996; 49(11): 1215-1219.

[50] Redelmeier, DA , Goldstein, RS, Min, ST *et al*. Spirometry and Dyspnea in Patients With COPD. When Small Differences Mean Little. *Chest* 1996; 109(5): 1163-1168.

[51] Redelmeier, DA, Guyatt, GH and Goldstein, RS. On the Debate over Methods for Estimating the Clinically Important Difference. *Journal of Clinical Epidemiology* 1996; 49(11): 1223-1224.

[52] Middel, B. Statistical significant change versus relevant or important change in (quasi) experimental design: Some conceptual and methodological problems in estimating magnitude of intervention-related change in health services research. *International Journal of Integrated Care* 2002; 2: e15.

[53] Cohen, J. A power primer. *Psychological Bulletin* 1992; 112(1): 155-159.

[54] Wyrwich, KW, Nienaber, NA, Tierney, WM *et al.* Linking Clinical Relevance and Statistical Significance in Evaluating Intra-Individual Changes in Health-Related Quality of Life. *Medical Care* 1999; 37(5): 469-478.

[55] Wyrwich, KW, Tierney, WM and Wolinsky, FD. Further Evidence Supporting an SEM-Based Criterion for Identifying Meaningful Intra-Individual Changes in Health-Related Quality of Life. *Journal of Clinical Epidemiology* 1999; 52(9): 861-873.

[56] Jacobsen, NS and Truax, P. Clinical Significance: A Statistical Approach to Defining Meaningful Change in Psychotherapy Research. *Journal of Consulting and Clinical Psychology* 1991; 59(1): 12-19.

[57] Lassere, MN, Van der Heijde, D and Johnson, KR. Foundations of the Minimal Clinically Important Difference for Imaging. *The Journal of Rheumatology* 2001; 28(4): 890-891.

[58] Wyrwich, KW, Fihn, SD, Tierney, WM *et al.* Clinically Important Changes in Health-related Quality of Life for Patients with Chronic Obstructive Pulmonary Disease. An Expert Consensus Panel Report. *Journal of General Internal Medicine* 2003; 18(3): 196-202.

[59] Grovle, L, Haugen, AJ, Hasvik, E *et al.* Patients ratings of global perceived change during 2 years were strongly influenced by the current health status. *Journal of Clinical Epidemiology* 2014; 67(5): 508-515.

[60] Jaeschke, RJ, Guyatt, GH, Keller, J *et al.* Interpreting Changes in Quality of Life Score in N of 1 Randomized Trials. *Controlled Clinical Trials* 1991; 12(4 Suppl): 226S-233S.

[61] Sundh, J, Montgomery, S, Hasselgren, M *et al.* Change in health status in COPD: a seven-year follow-up cohort study. *npj Primary Care Respiratory Medicine* 2016; 26: 16073.

[62] Nagai, K, Makita, H, Suzuki, M *et al.* Differential changes in quality of life components over 5 years in chronic obstructive pulmonary disease patients. *International Journal of COPD* 2015; 10: 745-757.

[63] Hajiro, TH and Nishimura, K. Minimal clinically significant difference in health status: the thorny path of health status measures? *European Respiratory Journal* 2002; 19(3): 390-391.

[64] Jones, PW, Beeh, KM, Chapman, KR *et al.* Minimal Clinically Important Differences in Pharmacological Trials. *American Journal of Respiratory Critical Care Medicine* 2014; 189(3): 250-255.

[65] Wacker, ME, Jörres, RA, Karch, A *et al.* Assessing health-related quality of life in COPD: comparing generic and disease-specific instruments with focus on comorbidities. *BMC Pulmonary Medicine* 2016; 16(1): 70.

[66] Wright, JG. The Minimal Important Difference: Who's to Say What Is Important? *Journal of Clinical Epidemiology* 1996; 49(11): 1221-1222.

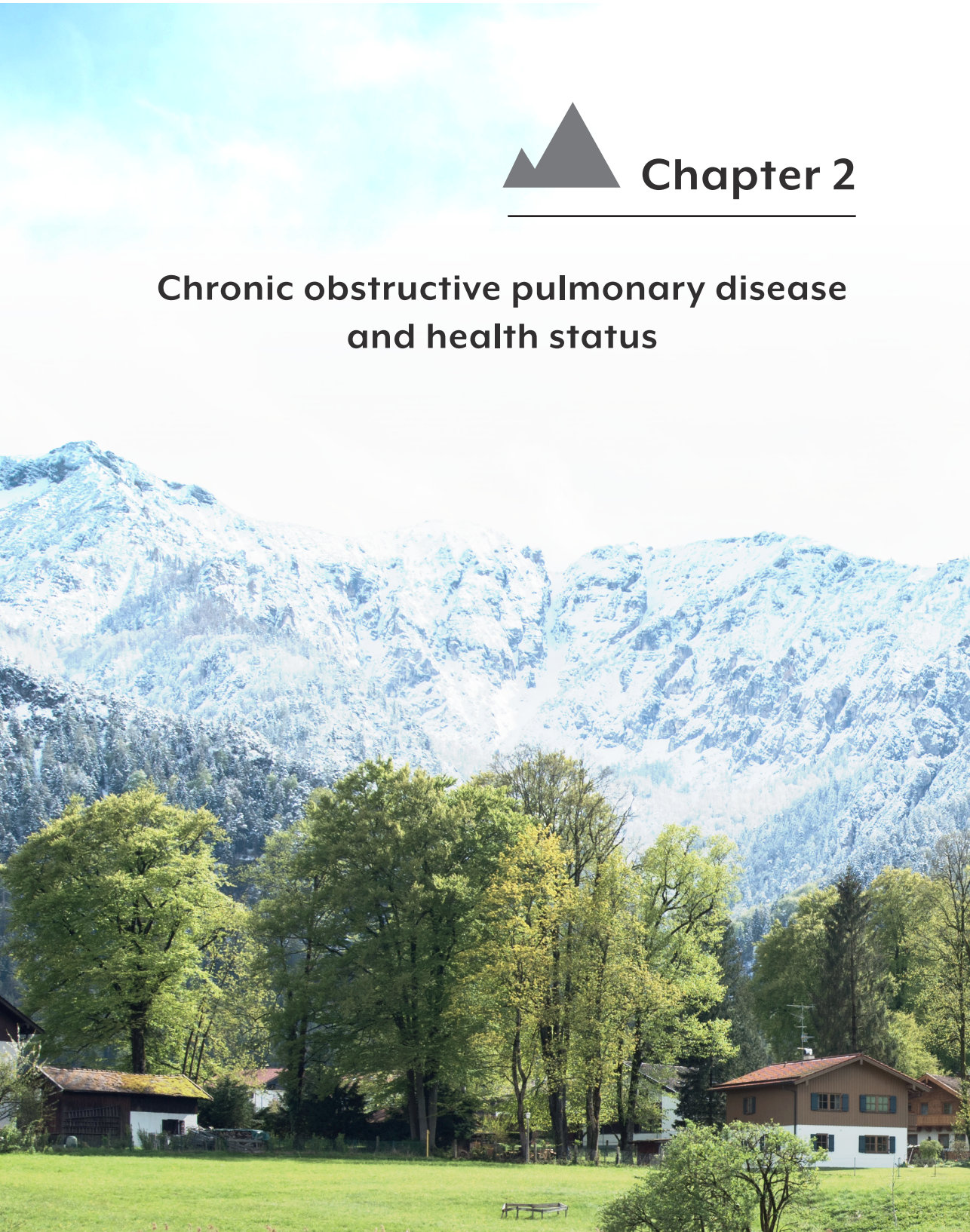
[67] Yoo, JY, Kim, YS, Kim, SS *et al.* Factors affecting the trajectory of health-related quality of life in COPD patients. *International Journal of Tuberculosis and Lung Disease* 2015; 20(6): 738-746.

[68] Martinez, CH, Diaz, AA, Parulekar, AD *et al.* Age-Related Differences in Health-Related Quality of Life in COPD. An Analysis of the COPDGene and SPIROMICS Cohorts 2016. *Chest* 2016; 149(4): 927-935.

[69] Rubinsztajn, R, Przybylowski, T, Maskey-Warzechowska, M *et al.* Exacerbations of Chronic Obstructive Pulmonary Disease and Quality of Life of Patients. *Advances in Experimental Medicine, Biology - Neuroscience and Respiration* 2016; 884: 69-74.

[70] Kim, JK, Lee, SH, Lee, BH *et al.* Factors associated with exacerbation in mild-to-moderate COPD patients. *International Journal of COPD* 2016; 11: 1327-1333.

[71] Schultz, K, Jelusic, D, Wittmann, M *et al.* Inspiratory muscle training does not improve clinical outcomes in 3-week COPD rehabilitation: results from a randomised controlled trial. *European Respiratory Journal* 2018; 51(1): 1702000.



Chapter 2

Chronic obstructive pulmonary disease and health status

2.1 Introduction and rationale

The previous chapter outlined the rationale for this thesis and highlighted the need to investigate the dynamics of the minimal clinically important difference (MCID) of health status instruments as well as the known problems with this parameter. Specifically, this thesis will focus on the various health status instruments and MCIDs used for patients with chronic obstructive pulmonary disease (COPD), a leading cause of morbidity and mortality worldwide. In *Chapter I (Box I)*, a representative case of COPD during pulmonary rehabilitation (PR) was presented, detailing the impact of the disease on that patient's life and well-being. This case emphasises the major discrepancy between the attainment of objective physical outcomes (e.g., spirometry) and the experienced burden of disease symptoms and daily functional or mental limitations specific to COPD. In this chapter, more detail will be provided about the background of this chronic disease and the measurement of health status in this population.

2.2 Chronic obstructive pulmonary disease

2.2.1 Definition

COPD has been defined as "*a chronic respiratory disorder that is characterised by persistent respiratory symptoms and obstructive airflow limitation*" [1-2]. Moreover, the respiratory symptoms and airflow obstruction are not fully reversible and are mostly progressive in nature. It is an under-diagnosed, life-threatening lung disease.

2.2.2 Pathophysiology

COPD results from a complex interaction between an individual's genes and his or her environment [1]. In the lungs of affected patients, there is a chronic abnormal inflammatory response to noxious particles and gases that results in typical pathological changes [2-3]. The most important environmental cause of COPD is tobacco smoke, but other situational and personal factors are known to be important, including atmospheric pollution, biomass fuels, occupational exposure, age, gender and a lower socio-economic status [1-3]. Alpha-1-antitrypsin deficiency may be a cause of COPD too.

The abnormal classification of the immune response in COPD concerns the fact that the observed inflammation does not occur to a comparable extent in healthy individuals. The presence of an extensive, chronic, innate, and adaptive inflammatory response interferes with normal repair and defence mechanisms in the lungs. This promotes and causes *small airway fibrosis and obstruction* [1-3], inducing parenchymal tissue destruction that can result in *emphysema*. In the process of airway fibrosis and obstruction, the abnormal

inflammatory response disrupts the epithelial barrier and interferes with the mucociliary clearance apparatus. This leads to an accumulation of inflammatory mucous exudates in the small airway lumen, because of mucus hypersecretion and ciliary dysfunction [3]. Inflammatory cells then infiltrate the airway walls and cause damage, whereupon the deposition of connective tissue leads to remodelling and wall thickening. Consequently, the lumen is reduced and restricts the normal increase in diameter during lung inflation. In addition, an imbalanced proteinase/anti-proteinase relationship and the presence of oxidative stress each contribute to the pathophysiology [1-3]. Overall, the abnormal processes in COPD are progressive and result in increased resistance and narrowing of the small airways, with emphysematous destruction leading to increased lung compliance due to diminished elastic recoil [3]. This creates a prolonged time for lung emptying that is defined clinically as obstructive airflow limitation.

2.2.3 Epidemiology

Despite being preventable and treatable, COPD is the leading cause of morbidity and mortality worldwide after ischaemic heart disease and stroke [2, 4-7]. As such it creates substantial economic and social burdens that are increasing [1, 8]. Current prevalence estimates range widely from less than 6% to over 19% [1]. The worldwide prevalence in adults older than 40 years is 10%–11%, equivalent to 384 million cases of COPD in 2010 [1, 4, 9-10]. Worldwide, 2.9 million deaths were recorded due to COPD in 2010 [6-7], with most of these occurring in low-income regions such as Asia and central Africa [11]. However, its morbidity statistics are arguably most notable, it has been reported to cause a staggering 76 million disability adjusted life years (DALYs) [12] and over 29 million years lived with disability (YLD) [13], ranking ninth and fifth, respectively.

The burden of COPD increases gradually from primary to secondary and tertiary care [14]. It has been estimated to account for approximately 6% of the total health care budget in the European Union (EU), equivalent to ≥50% of the costs for all respiratory diseases [1]. Exacerbations and limited work productivity account for most of this burden [1]. Here, an exacerbation is defined as *"a worsening of respiratory symptoms beyond normal day to day variation that requires additional medication or a change in therapy"* [1]. The burden of disease is often worse, because COPD frequently exists with cardiovascular, respiratory, metabolic, osteo-skeletal, and gastro-intestinal comorbidities (*Table I*) [1, 15-18]. Depression and/or anxiety are also common in COPD, with a mean prevalence of 27% (range 15.2%–37.5%) [19].

2.2.4 Symptoms

Fibrosis and obstruction of the small airways, coupled with emphysematous destruction of the airway walls, leads to obstructive airflow limitation, hyperinflation, air trapping, abnormal gas exchange, mucus hypersecretion, ciliary dysfunction and/or pulmonary hypertension [1-3]. The primary symptoms of COPD that result from this pathology are progressive dyspnoea, breathlessness, chronic cough and/or chronic sputum production [1, 18, 20], with wheezing and chest tightness sometimes present [1]. Symptoms that exist secondary to these include anxiety, panic, fear, frustration, and fatigue [20]. Fatigue may also result from the significant reduction in physical activity associated with COPD [21]. Furthermore, extra-pulmonary or systemic effects are not uncommon, including general signs of systemic inflammation, oxidative stress, and activated inflammatory cells; skeletal muscle dysfunction including cachexia and exercise limitations; nutritional abnormalities and weight loss due to malnutrition and cachexia; and cardiovascular, nervous and osteo-skeletal system effects [1-2, 22]

Table 1: Frequently observed comorbidities in patients with COPD

Cardiovascular	Cardiac dysrhythmias Coronary artery disease Heart failure Hypertension Peripheral artery disease Pulmonary hypertension
Metabolic	Cachexia Diabetes mellitus type 2 Hyperlipidemia / hypertriglyceridemia
Osteo-skeletal	Osteoporosis Skeletal muscle dysfunction
Mental	Anxiety Cognitive impairment Depression
Gastro-intestinal	Gastric / Duodenal ulcus Helicobacter Pylori infections Reflux
Respiratory	Asthma / Asthma-COPD overlap syndrome (ACOS) Bronchiectasis Lung cancer Obstructive sleep apnoea syndrome (OSAS) Pulmonary fibrosis

Abbreviations: ACOS, asthma-COPD overlap syndrome; COPD, chronic obstructive pulmonary disease; OSAS, obstructive sleep apnoea syndrome.

2.2.5 Diagnostic considerations

To confirm the diagnosis COPD, spirometry is traditionally required with a post-bronchodilator forced expiratory volume in one second (FEV_1) / forced vital capacity (FVC) ratio of ≤0.70 or ≤70% [1-2, 23]. The predicted FEV_1 percentage determines the degree of airway obstruction and the severity of the COPD, giving traditional classification categories I–IV of the global initiative for obstructive lung disease (GOLD) (*Table 2*). Nevertheless, identifying patients with COPD is difficult, with asthma being a particularly common misdiagnosis. Although demonstrating reversible airway obstruction during post-bronchodilation spirometry should confirm asthma [1], this does not account for those patients with a diagnostic overlap of asthma and COPD (ACOS). Moreover, most patients with COPD remain undiagnosed, because they are frequently asymptomatic [25], especially if they have only mild to moderate disease. Therefore, in many cases, an exacerbation is frequently the first presentation of symptomatic COPD [26]. Given that there is a more rapid decline in FEV_1 when COPD is less severe, early diagnosis and treatment is of major importance [25].

Table 2: Traditional spirometry-based classification of COPD according to GOLD grade I-IV (Adapted from NHG Standaard COPD 2015 [24] and GOLD 2017 [1])

Degree of airway obstruction	FEV_1/FVC	FEV_1 (% predicted)
I Mild	≤0.70	≥ 80
II Moderate		≤50 to <80
III Severe		≤30 to <50
IV Very severe		< 30

Abbreviations: COPD, chronic obstructive pulmonary disease; FEV_1, forced expiratory volume in one second; FVC, forced vital capacity; GOLD, global initiative for obstructive lung disease; NHG, Nederlands Huisartsen Genootschap.

2.2.6 Therapy

There is no cure for COPD. However, behavioural modifications and early treatment may improve symptoms and slow disease progression [25]. International guidelines recommend smoking cessation aids, exercise and physiotherapy, self-management and education, pharmacotherapy (e.g., short- and long-acting bronchodilators, glucocorticoids, theophylline, and/or antibiotics for exacerbations), long-term oxygen therapy or ventilatory support, PR, nutritional support and/or surgical interventions (e.g., lung volume reduction, bronchoscopic coiling, bullectomy, or lung transplantation) [1-2, 23, 25-26]. It is therefore essential that practitioners focus on planned COPD care to prevent exacerbations, slow disease progression, and reduce the need for rescue therapy of augmented symptoms [26]. One should also remember that patient's expectations and needs affect treatment adherence and outcomes, with patients often reasonably

wanting to have fewer exacerbations and fewer symptoms of dyspnoea, cough, and sputum production [27].

PR is an evidence-based, multidisciplinary, and comprehensive intervention for COPD, especially beneficial for symptomatic patients with decreased lung function [1, 28]. Programmes are individualised and include detailed patient assessment, exercise training, education and psychosocial support [28-30]. They are designed to reduce symptoms, optimise functional status, increase participation, and minimise health care costs by stabilising the disease [28-29]. PR can significantly improve dyspnoea, quality of life (QoL), and psychosocial well-being, and can decrease health care utilisation in a cost-effective manner [28, 30-36]. Of note, increasing physical activity – defined as *bodily movement produced by skeletal muscles that results in energy expenditure* – may have favourable effects on lung function decline, FEV_1 levels, COPD symptoms, QoL, exacerbations, and all-cause mortality [37].

The patient interview in *Chapter 1* (*Box 1*) highlighted a case of PR in which add-on inspiratory muscle training (IMT) was used as part of a randomised controlled clinical trial (RCT). Although the available evidence is contentious, IMT could be an additional component of PR [38-42]. COPD results in patients having significant inspiratory muscle weakness that may contribute to dyspnoea and exercise intolerance [39]. By implementing resistance training during inspiration, IMT may improve a patient's capacity for higher ventilation levels, thereby reducing the sensation of dyspnoea [38]. IMT is, however, not officially recommended at present [28].

2.3 Health status in patients with COPD

2.3.1 Rationale for measuring health status

Airflow limitation, as measured by the FEV_1, is important in diagnosing and measuring COPD. However, the clinical features of this disease are much more heterogeneous and cannot be captured by the FEV_1 alone [43]. Indeed, there is only a weak to moderate correlation between spirometry results, symptoms, and experienced QoL impairment [1, 20, 44-50]. COPD results in worse health-related quality of life (HRQoL), greater impairment of work productivity, and greater health care utilisation [51]. Health status, as the standardised measure of HRQoL, seems to deteriorate over time in these patients, though analysis in a 5-year study indicated that accurate therapy could produce improvements [52-53]. Several factors are known to predict the worsening health status in COPD, including increased dyspnoea symptoms, depression and anxiety, functional status deficits, exacerbations and hospital admissions [52, 54]. A more severe health status has

also been associated with increased age, sleep disturbances, depression, worse COPD symptoms, and frequent exacerbations (>2 in the previous year) [55]. The resulting worse HRQoL then presents with higher morbidity and mortality [56]. COPD exacerbations are among the most important factors associated with decreased QoL [57-58], with higher GOLD I–IV grades and female sex associated with a higher prevalence of exacerbations [57-58]. In the US, the risk of in-hospital death due to an acute exacerbation of COPD is 11% [59].

2.3.2 Integration of health status in COPD assessment

The optimal care for patients with COPD requires an individualised approach that recognises all aspects of the disease and commitment from all stakeholders [60]. In 2017, the GOLD strategic update proposed an ABCD framework to help deliver more comprehensive COPD assessment [1]. In addition to the spirometry-derived GOLD classification (*Table 2*), this framework assessed the exacerbation history (≥2 or <2 exacerbations in the past year) and the symptoms measured by specific health status instruments (*Figure 1*) [61]. The instruments used to evaluate symptoms in this revised framework are the COPD Assessment Test (CAT) and the modified Medical Research Council dyspnoea scale (mMRC). Cutoff values have been defined for both the CAT (10 points) and the mMRC (2 points) to distinguish symptomatic from asymptomatic or less symptomatic patients [1]. However, there remains debate as to what extent the established cutoff values may lead to misclassification and discrepancies [1, 62-70]. Corresponding values for symptomatic patients have been defined for other frequently used health status tools for COPD too, including the Clinical COPD Questionnaire (CCQ, 1.5-2.0 points) and the St. George's Respiratory Questionnaire (SGRQ, 20-25 points) [64, 68]. Based on the GOLD ABCD framework, a risk classification and pharmacological treatment algorithms have been developed.

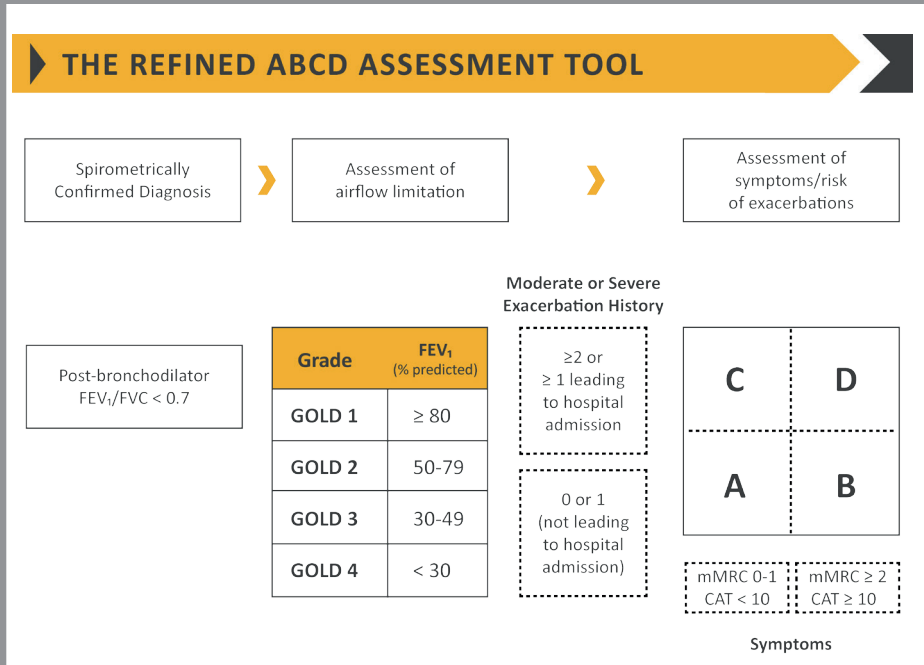
2.3.3 Recommended health status tools

Health status patient-reported outcomes (PROs) can quantify the extent to which the physiological effects and symptoms of COPD affect a patient's health and function [49]. They include the major concerns for patients with COPD, such as breathlessness, dyspnoea, fatigue, cough, sputum production, physical function and exercise tolerance, social function, depression and/or anxiety and exacerbations [45, 49, 71].

Many health status instruments and functional status tools have been developed for use by patients with COPD (*Table 3*). Certain tools have been used more frequently than others, and at present, the CAT and CCQ are recommended in clinical practice [1, 72]. The CAT is an 8-item unidimensional questionnaire that includes questions about cough, phlegm, chest tightness, breathlessness, walking up stairs/hills, activity limitation

at home, sleep, confidence leaving home, and energy (*Supplementary material 2.4.1*) [73]. Each item is scored on a scale from 0 to 5 points, totalling a maximum of 40 points. The CAT has been shown to be a reliable, valid, reproducible, and responsive tool with strong discriminative properties [66, 73-84]. The CCQ is a 10-item multi-dimensional health status tool comprising three domains including symptoms (4 items), functional status (4 items), and mental state (2 items) (*Supplementary material 2.4.2*) [85]. Items are scored on a scale from 0 to 6 points. The total and domain scores are determined by adding the relevant item scores and dividing this by the number of items. The CCQ also has strong psychometric and discriminative properties [85-89]. Higher scores indicate worse HRQoL on both the CAT and CCQ, and both tools are considered equally reliable, valid, and reproducible [90]. Moreover, the correlation is high between the CAT and CCQ, and it has been suggested that they could be used interchangeably [91].

Figure 1: GOLD ABCD Framework
(Published in Pocket Guide to COPD Diagnosis, Management, and Prevention: a Guide for Health Care Professionals 2017 [61], printed with permission granted by the GOLD committee)



Abbreviations: CAT, COPD Assessment Test; COPD, chronic obstructive pulmonary disease; FEV_1, forced expiratory volume in one second; FVC, forced vital capacity; GOLD, global initiative for obstructive lung disease; mMRC, modified Medical Research Council dyspnoea scale; RCP, routine clinical practice.

Table 3: Overview of the variety of COPD health status, functional status and dyspnoea instruments

COPD-specific health status instruments	Generic health status instruments
Dyspnoea/breathlessness/cough instruments	
Baseline/Transitional Dyspnoea Index (BDI/TDI) [100-102]	Dartmouth Northern New England Primary Care Cooperative Information Project chart system (DartmCoop) [105]
Borg Scale [56]	EuroQol-5D (EQ-5D) [56, 104, 110]
Breathing Problems Questionnaire (BPQ) [48]	Hyland Scale [105]
Breathlessness Cough and Sputum Scale (BCSS) [103]	Linear Analogue Scale/Visual Analogue Scale (LAS/VAS-8) [105]
Cough Severity Diary (CSD) [49, 104]	Measure Yourself Medical Outcome Profile (MYMOP) [105]
Dyspnoea-12 Tool [49]	Nottingham Health Profile (NHP) [105]
Dyspnoea Questionnaire Computer Adaptive Test (DMQ-CAT) [49]	Quality of Well Being Self-Administered (QWBSA) [105]
Global Chest Symptoms Questionnaire (GCSQ) [49]	Short-Form-12 (SF-12) [56, 104]
Leicester Cough Questionnaire (LCQ) [49, 104]	Short-Form 36 (SF-36) [56, 104]
Modified Medical Research Council Dyspnoea Scale (mMRC) [56, 99]	Sickness Impact Profile (SIP) [105]
Maugeri Respiratory Failure (MRF-28) [105]	World Health Organisation Quality of Life short version list (WHOQOLBREF) [105]
Severe Respiratory Insufficiency Questionnaire (SRI) [105]	
Shortness of Breath with Daily Activities Questionnaire (SOBDA) [49]	
University of California San Diego Shortness of Breath Questionnaire (UCSD-SOBQ) [49]	
	Functional status physical tools
Multi-domain instruments	6 Minute Walking Distance (6MWD) [48-49, 56, 111-112, 104]
Airways Questionnaire (AQ) [104]	Activity monitors (e.g. step counters) [48-49, 56, 111-112, 104]
Chronic Respiratory Questionnaire (CRQ) [106]	PROActive tools [49]
COPD Assessment Test (CAT) [73]	Shuttle Walking Test (SWT) [48-49, 56, 111-112, 104]
Clinical COPD Questionnaire (CCQ) [85]	
COPD Specific Item Bank (COPD-SIB) [107]	**Functional status questionnaires**
Health States COPD (HS-COPD) [104]	Capacity of Daily Living during the Morning Questionnaire (CDLM) [49]
Living with COPD Questionnaire (LCOPD) [49]	London Chest Activity of Daily Living Questionnaire (LCADL) [49, 113]
Mc Gill COPD Quality of Life Questionnaire [103]	Pulmonary Functional Status and Dyspnoea Questionnaire (PFSDQ) [48-49, 56, 111-112, 104]
Quality of Life for Respiratory Illness Questionnaire (QOLRIQ) [48, 108]	Pulmonary Functional Status Scale (PFSS) [48-49, 56, 111-112, 104]
Respiratory Quality of Life Questionnaire (RQLQ) [105]	
St. George's Respiratory Questionnaire (SGRQ) [93-94]	
Visual Simplified Respiratory Questionnaire (VSRQ) [105, 109]	
Fatigue/energy instruments	
Manchester COPD Fatigue Scale [49, 104]	
Work-related instruments	
Work Productivity and Activity Impairment Questionnaire (WPAI-COPD) [49, 103]	

Note: The provided lists are not intended to be comprehensive. The (multi-domain) COPD-specific health status instruments highlighted in bold italics will form the basis of analysis in this thesis.

Although other questionnaires are well established, such as the SGRQ and the Chronic Respiratory Questionnaire (CRQ), they are too long for use in daily practice. Indeed, the CAT, CCQ and SGRQ correlated well, but the CAT and CCQ benefit from being faster and easier to administer [72, 90, 92]. The SGRQ is a self-completed standardised multi-

dimensional HRQoL instrument that contains 50 items (*Supplementary material 2.4.3*) [93-94]. It includes domains for symptoms, activity, and impact, with each item having an empirically derived weight in the scoring algorithm. Domain and total scores range from 0 (best health status) to 100 (worst health status), and the results are considered valid, reliable, repeatable, responsive and correlated with reference measures [93-96]. Both the SGRQ and CRQ had similar reliability, validity, and responsiveness in patients with COPD, with no clear evidence in preference of one instrument over the other [97-98]. In addition to these health status questionnaires, the (m)MRC dyspnoea scale is recommended by GOLD to measure activity limitations due to dyspnoea [1, 99].

In the current thesis, the focus will be on the MCIDs of the (multi-domain) COPD-specific CAT, CCQ, and SGRQ health status instruments (*Table 3*).

2.4 Supplementary material

2.4.1 The COPD Assessment Test (CAT) (printed with permission)

Your name:

Today's date:

CAT
COPD Assessment Test

How is your COPD? Take the COPD Assessment Test™ (CAT)

This questionnaire will help you and your healthcare professional measure the impact COPD (Chronic Obstructive Pulmonary Disease) is having on your wellbeing and daily life. Your answers, and test score, can be used by you and your healthcare professional to help improve the management of your COPD and get the greatest benefit from treatment.

For each item below, place a mark (X) in the box that best describes you currently. Be sure to only select one response for each question.

Example: I am very happy 0 X 2 3 4 5 I am very sad

			SCORE
I never cough	0 1 2 3 4 5	**I cough all the time**	
I have no phlegm (mucus) in my chest at all	0 1 2 3 4 5	**My chest is completely full of phlegm (mucus)**	
My chest does not feel tight at all	0 1 2 3 4 5	**My chest feels very tight**	
When I walk up a hill or one flight of stairs I am not breathless	0 1 2 3 4 5	**When I walk up a hill or one flight of stairs I am very breathless**	
I am not limited doing any activities at home	0 1 2 3 4 5	**I am very limited doing activities at home**	
I am confident leaving my home despite my lung condition	0 1 2 3 4 5	**I am not at all confident leaving my home because of my lung condition**	
I sleep soundly	0 1 2 3 4 5	**I don't sleep soundly because of my lung condition**	
I have lots of energy	0 1 2 3 4 5	**I have no energy at all**	
		TOTAL SCORE	

2.4.2 The Clinical COPD Questionnaire (CCQ) (printed with permission)

Patient number:_____
Date:_____

CLINICAL COPD QUESTIONNAIRE

Please **circle** the number of the response that best describes how you have been feeling during the **past week**.
(Only **one** response for each question).

On average, **during the past week**, how often did you feel:	never	hardly ever	a few times	several times	many times	a great many times	almost all the time
1. Short of breath **at rest**?	0	1	2	3	4	5	6
2. Short of breath **doing physical activities**?	0	1	2	3	4	5	6
3. **Concerned** about getting a cold or your breathing getting worse?	0	1	2	3	4	5	6
4. **Depressed (down)** because of your breathing problems?	0	1	2	3	4	5	6
In general, **during the past week**, how much of the time:							
5. Did you **cough**?	0	1	2	3	4	5	6
6. Did you **produce phlegm**?	0	1	2	3	4	5	6
On average, **during the past week**, how limited were you in these activities **because of your breathing problems**:	not limited at all	very slightly limited	slightly limited	moderately limited	very limited	extremely limited	totally limited /or unable to do
7. **Strenuous physical activities** (such as climbing stairs, hurrying, doing sports)?	0	1	2	3	4	5	6
8. **Moderate physical activities** (such as walking, housework, carrying things)?	0	1	2	3	4	5	6
9. **Daily activities at home** (such as dressing, washing yourself)?	0	1	2	3	4	5	6
10. **Social activities** (such as talking, being with children, visiting friends/ relatives)?	0	1	2	3	4	5	6

2.4.3 The St. George's Respiratory Questionnaire (SGRQ) (printed with permission)

ST. GEORGE'S RESPIRATORY QUESTIONNAIRE ORIGINAL ENGLISH VERSION

ST. GEORGE'S RESPIRATORY QUESTIONNAIRE (SGRQ)

This questionnaire is designed to help us learn much more about how your breathing is troubling you and how it affects your life. We are using it to find out which aspects of your illness cause you most problems, rather than what the doctors and nurses think your problems are.

Please read the instructions carefully and ask if you do not understand anything. Do not spend too long deciding about your answers.

Before completing the rest of the questionnaire:

Please tick in one box to show how you describe your current health:

Very good	Good	Fair	Poor	Very poor
☐	☐	☐	☐	☐

P.W. Jones, PhD FRCP
Professor of Respiratory Medicine,
St. George's University of London,
Jenner Wing,
Cranmer Terrace,
London SW17 ORE, UK.

Tel. +44 (0) 20 8725 5371
Fax +44 (0) 20 8725 5955

continued…

St. George's Respiratory Questionnaire
PART 1

Questions about how much chest trouble you have had over the past 3 months.

Please tick (✓) *one box for each question*:

	most days a week	several days a week	a few days a month	only with chest infections	not at all
1. Over the past 3 months, I have coughed:	☐	☐	☐	☐	☐
2. Over the past 3 months, I have brought up phlegm (sputum):	☐	☐	☐	☐	☐
3. Over the past 3 months, I have had shortness of breath:	☐	☐	☐	☐	☐
4. Over the past 3 months, I have had attacks of wheezing:	☐	☐	☐	☐	☐

5. During the past 3 months how many severe or very unpleasant attacks of chest trouble have you had?

Please tick (✓) *one*:

- more than 3 attacks ☐
- 3 attacks ☐
- 2 attacks ☐
- 1 attack ☐
- no attacks ☐

6. How long did the worst attack of chest trouble last?
(Go to question 7 if you had no severe attacks)

Please tick (✓) *one*:

- a week or more ☐
- 3 or more days ☐
- 1 or 2 days ☐
- less than a day ☐

7. Over the past 3 months, in an average week, how many good days (with little chest trouble) have you had?

Please tick (✓) *one*:

- No good days ☐
- 1 or 2 good days ☐
- 3 or 4 good days ☐
- nearly every day is good ☐
- every day is good ☐

8. If you have a wheeze, is it worse in the morning?

Please tick (✓) *one:*

- No ☐
- Yes ☐

St. George's Respiratory Questionnaire
PART 2

Section 1

How would you describe your chest condition?

Please tick (✓) *one*:

- The most important problem I have ☐
- Causes me quite a lot of problems ☐
- Causes me a few problems ☐
- Causes no problem ☐

If you have ever had paid employment.

Please tick (✓) *one*:

- My chest trouble made me stop work altogether ☐
- My chest trouble interferes with my work or made me change my work ☐
- My chest trouble does not affect my work ☐

Section 2

Questions about what activities usually make you feel breathless <u>these days</u>.

Please tick (✓) in ***each box*** that applies to you ***these days***:

	True	False
Sitting or lying still	☐	☐
Getting washed or dressed	☐	☐
Walking around the home	☐	☐
Walking outside on the level	☐	☐
Walking up a flight of stairs	☐	☐
Walking up hills	☐	☐
Playing sports or games	☐	☐

continued…

St. George's Respiratory Questionnaire
PART 2

Section 3

Some more questions about your cough and breathlessness <u>these days</u>.

	Please tick (✓) in ***each box*** that applies to you ***these days***: True	False
My cough hurts	☐	☐
My cough makes me tired	☐	☐
I am breathless when I talk	☐	☐
I am breathless when I bend over	☐	☐
My cough or breathing disturbs my sleep	☐	☐
I get exhausted easily	☐	☐

Section 4

Questions about other effects that your chest trouble may have on you <u>these days</u>.

	Please tick (✓) in ***each box*** that applies to you ***these days***: True	False
My cough or breathing is embarrassing in public	☐	☐
My chest trouble is a nuisance to my family, friends or neighbours	☐	☐
I get afraid or panic when I cannot get my breath	☐	☐
I feel that I am not in control of my chest problem	☐	☐
I do not expect my chest to get any better	☐	☐
I have become frail or an invalid because of my chest	☐	☐
Exercise is not safe for me	☐	☐
Everything seems too much of an effort	☐	☐

Section 5

Questions about your medication, if you are receiving no medication go straight to section 6.

	Please tick (✓) in ***each box*** that applies to you ***these days***: True	False
My medication does not help me very much	☐	☐
I get embarrassed using my medication in public	☐	☐
I have unpleasant side effects from my medication	☐	☐
My medication interferes with my life a lot	☐	☐

continued…

f:\institut\cultadap\project\gsk1881\question\final versions\sgrqoriq.doc 14/03/03

St. George's Respiratory Questionnaire
PART 2

Section 6

These are questions about how your activities might be affected by your breathing.

Please tick (✓) in ***each box*** that applies to you ***because of your breathing***:

	True	False
I take a long time to get washed or dressed	☐	☐
I cannot take a bath or shower, or I take a long time	☐	☐
I walk slower than other people, or I stop for rests	☐	☐
Jobs such as housework take a long time, or I have to stop for rests	☐	☐
If I walk up one flight of stairs, I have to go slowly or stop	☐	☐
If I hurry or walk fast, I have to stop or slow down	☐	☐
My breathing makes it difficult to do things such as walk up hills, carrying things up stairs, light gardening such as weeding, dance, play bowls or play golf	☐	☐
My breathing makes it difficult to do things such as carry heavy loads, dig the garden or shovel snow, jog or walk at 5 miles per hour, play tennis or swim	☐	☐
My breathing makes it difficult to do things such as very heavy manual work, run, cycle, swim fast or play competitive sports	☐	☐

Section 7

We would like to know how your chest <u>usually</u> affects your daily life.

Please tick (✓) in ***each box*** that applies to you ***because of your chest trouble***:

	True	False
I cannot play sports or games	☐	☐
I cannot go out for entertainment or recreation	☐	☐
I cannot go out of the house to do the shopping	☐	☐
I cannot do housework	☐	☐
I cannot move far from my bed or chair	☐	☐

continued…

St. George's Respiratory Questionnaire

Here is a list of other activities that your chest trouble may prevent you doing. (You do not have to tick these, they are just to remind you of ways in which your breathlessness may affect you):

Going for walks or walking the dog

Doing things at home or in the garden

Sexual intercourse

Going out to church, pub, club or place of entertainment

Going out in bad weather or into smoky rooms

Visiting family or friends or playing with children

Please write in any other important activities that your chest trouble may stop you doing:

. . . . . .

. . . . . .

. . . . . .

. . . . . .

Now would you tick in the box (one only) which you think best describes how your chest affects you:

It does not stop me doing anything I would like to do ☐

It stops me doing one or two things I would like to do ☐

It stops me doing most of the things I would like to do ☐

It stops me doing everything I would like to do ☐

Thank you for filling in this questionnaire. Before you finish would you please check to see that you have answered all the questions.

UK/ English (original) version 6

f:\institut\cultadap\project\gsk1881\question\final versions\sgrqoriq.doc 14/03/03

2.5 References

[1] Global initiative for Chronic Obstructive Lung Disease (GOLD). Global Strategy for the Diagnosis, Management and Prevention of COPD (2017 Report) 2017; Available at http://www.goldcopd.org/gold-2017-global-strategy-diagnosis-management-prevention-copd/. Accessed 9th of July 2018.

[2] Celli, BR, MacNee, W, ATS/ERS Task Force. Standards for the diagnosis and treatment of patients with COPD: a summary of the ATS/ERS position paper. *European Respiratory Journal* 2004; 23(6): 932-946.

[3] Hogg, JC. Pathophysiology of airflow limitation in chronic obstructive pulmonary disease. *Lancet* 2004; 364(9435): 709-721.

[4] Adeloye, D, Chua, S, Lee, C *et al*. Global and regional estimates of COPD prevalence: Systematic review and meta-analysis. *Journal of Global Health* 2015; 5(2): 020415.

[5] US Burden of Disease Collaborators. The State of US Health, 1990-2010: Burden of Diseases, Injuries and Risk Factors. *Journal of the American Medical Association (JAMA)* 2013; 310(6): 591-608.

[6] Lozano, R, Naghavi, M, Foreman, K *et al*. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet* 2012; 380(9859): 2095-2128.

[7] Murray, CJL *et al*. Global, regional, and national age-sex specific mortality for 264 causes of death, 1980–2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet* 2017; 390(10100): 1151-1210.

[8] Lopez-Campos, JL, Wan, T and Soriano, JB. Global burden of COPD. *Respirology* 2016; 21(1): 14-23.

[9] Buist, AS, McBurnie, MA, Vollmer, WM *et al*. International variation in the prevalence of COPD (The BOLD Study): a population-based prevalence study. *Lancet* 2017; 370(9589): 741-750.

[10] Halbert, RJ, Natoli, JL, Gano, A *et al*. Global burden of COPD: systematic review and meta-analysis. *European Respiratory Journal* 2006(3); 28: 523-532.

[11] Burney, PGJ, Patel, J, Newson, R *et al*. Global and regional trends in COPD mortality, 1990–2010. *European Respiratory Journal* 2015; 45(5): 1239-1247.

[12] Murray, CJL, Vos, T, Lozano, R *et al*. Disability-adjusted Life Years (DALYs) for 291 diseases and injuries in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet* 2012; 380(9859): 2197-2223.

[13] Vos, T, Flaxman, AD, Naghavi, M *et al*. Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet* 2012; 380(9859): 2163-2196.

[14] Smid, DE, Spruit, MA, Houben-Wilke, S *et al*. Burden of COPD in patients treated in different care settings in the Netherlands. *Respiratory Medicine* 2016; 118: 76-83.

[15] Negewo, A, Gibson, PG and McDonald, VM. COPD and its comorbidities: Impact, measurement and mechanisms. *Respirology* 2015; 20(8): 1160-1171.

[16] Corlateanu, A, Covantev, S, Mathioudakis, AG *et al*. Prevalence and burden of comorbidities in Chronic Obstructive Pulmonary Disease. *Respiratory Investigation* 2016; 54(6): 387-396.

[17] Divo, MJ, Casanova, C, Marin, JM *et al*. Chronic obstructive pulmonary disease comorbidities network. *European Respiratory Journal* 2015; 46(3): 640-650.

[18] Worth, H, Buhl, R, Criée, C *et al*. The 'real-life' COPD patient in Germany: The DACCORD study. *Respiratory Medicine* 2016; 111: 64-71.

[19] Matte, DL, Pizzichini, MMM, Hoepers, ATC *et al*. Prevalence of depression in COPD: A systematic review and meta-analysis of controlled studies. *Respiratory Medicine* 2016; 117: 154-161.

[20] Barnett, M. Chronic obstructive pulmonary disease: a phenomenological study of patients' experiences. *Journal of Clinical Nursing* 2005; 14(7): 805-812.

[21] Anderson, WH, Wook Ha, J, Couper, DJ *et al*. Variability in objective and subjective measures affects baseline values in studies of patients with COPD. *PLoS ONE* 2017; 12(9): e0184606.

[22] Agusti, AGN, Noguera, A, Sauleda, J *et al*. Systemic effects of chronic obstructive pulmonary disease. *European Respiratory Journal* 2003; 21(2): 347-360.

[23] Miravitlles, M, Vogelmeier, C, Roche, N *et al*. A review of national guidelines for management of COPD in Europe. *European Respiratory Journal* 2016; 47(2): 625-637.

[24] Snoeck-Stroband, JB, Schermer, TRJ, Van Schayck, CP *et al*. NHG-Werkgroep Astma bij volwassenen en COPD. NHG-Standaard COPD (derde herziening). *Huisarts Wet* 2015; 58(4): 198-211.

[25] Price, D, Freeman, D, Cleland, J *et al*. Earlier diagnosis and earlier treatment of COPD in primary care. *Primary Care Respiratory Journal* 2011; 20(1): 15-22.

[26] Fromer, L. Implementing chronic care for COPD: planned visits, care coordination, and patient empowerment for improved outcomes. *International Journal of COPD* 2011; 6: 605-614.

[27] Braido, F, Baiardini, I, Molinengo, G *et al*. Choose your outcomes: From the mean to the personalized assessment of outcomes in COPD. An exploratory pragmatic survey. *European Journal of Internal Medicine* 2016; 34: 85-88.

[28] Ries, AL, Bauldoff, GS, Carlin, BW *et al*. Pulmonary Rehabilitation. Joint ACCP/AACVPR Evidence-Based Clinical Practice Guidelines. *Chest* 2007; 131(5 Suppl): 4S–42S.

[29] Vogiatzis, I and Zakynthinos, S. The physiological basis of rehabilitation in chronic heart and lung disease. *Journal of Applied Physiology* 2013; 115(1): 16-21.

[30] Spruit, MA, Singh, SJ, ZuWallack, R *et al*. An Official American Thoracic Society/European Respiratory Society Statement: Key Concepts and Advances in Pulmonary Rehabilitation. *American Journal of Respiratory Critical Care Medicine* Vol 2013; 188(8): e13-e64.

[31] Salman, GF, Mosier, MC, Beasley, BW *et al*. Rehabilitation for Patients with Chronic Obstructive Pulmonary Disease. Meta-analysis of Randomized Controlled Trials. *Journal of General Internal Medicine* 2003; 18(3): 213-221.

[32] Lacasse, Y, Goldstein, R, Lasserson, TJ *et al*. Pulmonary rehabilitation for chronic obstructive pulmonary disease (Review). *Cochrane Database of Systematic Reviews* 2006; 4: CD003793.

[33] Alfarroba, S, Rodrigues, F, Papoila, AL *et al*. Pulmonary Rehabilitation in COPD According to Global Initiative for Chronic Obstructive Lung Disease Categories. *Respiratory Care* 2016; 61(10): 1331-1340.

[34] Sahin, H, Varol, Y, Naz, I *et al*. The effect of pulmonary rehabilitation on COPD exacerbation frequency per year. *The Clinical Respiratory Journal* 2018; 12(1): 165-174.

[35] Rochester, CL, Vogiatzis, I, Holland, AE *et al*. An Official American Thoracic Society/European Respiratory Society Policy Statement: Enhancing Implementation, Use, and Delivery of Pulmonary Rehabilitation. *American Journal of Respiratory Critical Care Medicine* 2015; 192(11): 1373-1386.

[36] Meshe, OF, Claydon, LS, Bungay, H *et al*. The relationship between physical activity and health status in patients with chronic obstructive pulmonary disease following pulmonary rehabilitation. *Disability and Rehabilitation* 2017; 39(8): 746-756.

[37] Watz, H, Pitta, F, Rochester, CL *et al*. An official European Respiratory Society statement on physical activity in COPD. *European Respiratory Journal* 2014; 44(6): 1521-1537.

[38] Beaumont, M, Mialon, P, Le Ber-Moy, C *et al*. Inspiratory muscle training during pulmonary rehabilitation in chronic obstructive pulmonary disease: A randomized trial. *Chronic Respiratory Disease* 2015; 12(4): 305-312.

[39] Magadle, R, McConnell, AK, Beckerman, M *et al*. Inspiratory muscle training in pulmonary rehabilitation program in COPD patients. *Respiratory Medicine* 2007; 101(7): 1500-1505.

[40] Geddes, EL, O'Brien, K, Reid, WD *et al*. Inspiratory muscle training in adults with chronic obstructive pulmonary disease: An update of a systematic review. *Respiratory Medicine* 2008; 102(12): 1715-1729.

[41] O'Brien, K, Geddes, EL, Reid, DW *et al*. Inspiratory Muscle Training Compared With Other Rehabilitation Interventions in Chronic Obstructive Pulmonary Disease: a systematic review update. *Journal of Cardiopulmonary Rehabilitation and Prevention* 2008; 28(2): 128-141.

[42] Hill, K, Cecins, NM, Eastwood, PR *et al*. Inspiratory Muscle Training for Patients With Chronic Obstructive Pulmonary Disease: A Practical Guide for Clinicians. *Physical Medicine and Rehabilitation* 2010; 91(9): 1466-1470.

[43] Rennard, SI and Drummond, MB. Early chronic obstructive pulmonary disease: definition, assessment, and prevention. *Lancet* 2015; 385(9979): 1778-1788.

[44] Westwood, M, Bourbeau, J, Jones, PW *et al*. Relationship between FEV1 change and patient-reported outcomes in randomized trials of inhaled bronchodilators for stable COPD: a systematic review. *Respiratory Research* 2011; 12: 40.

[45] Tsiligianni, I, Kocks, J, Tzanakis, N *et al*. Factors that influence disease-specific quality of life or health status in patients with COPD: a review and meta-analysis of Pearson correlations. *Primary Care Respiratory Journal* 2011; 20(3); 257-268.

[46] Curtis, JR and Patrick, DL. The assessment of health status among patients with COPD. *European Respiratory Journal Supplement* 2003; 41: 36s-45s.

[47] Jones, PW. Issues Concerning Health-Related Quality of Life in COPD. *Chest* 1995; 107(5 Suppl): 187S-193S.

[48] Jones, PW. Health status measurement in chronic obstructive pulmonary disease. *Thorax* 2001; 56(11): 880-887.

[49] Jones, P, Miravittles, M, van der Molen, T *et al*. Beyond FEV1 in COPD: a review of patient-reported outcomes and their measurement. *International Journal of COPD* 2012; 7: 697-709.

[50] Martin, AL, Marvel, J, Fahrbach, K *et al*. The association of lung function and St. George's respiratory questionnaire with exacerbations in COPD: a systematic literature review and regression analysis. *Respiratory Research* 2016; 17: 40.

[51] Dhamane, AD, Witt, EA and Su, J *et al*. Associations Between COPD Severity and Work Productivity, Health-Related Quality of Life, and Health CareResource Use: A Cross-Sectional Analysis of National Survey Data. *Journal of Occupational and Environmental Medicine* 2016; 58(6): e191-197.

[52] Sundh, J, Montgomery, S, Hasselgren, M *et al*. Change in health status in COPD: a seven-year follow-up cohort study. *npj Primary Care Respiratory Medicine* 2016; 26: 16073.

[53] Nagai, K, Makita, H, Suzuki, M *et al*. Differential changes in quality of life components over 5 years in chronic obstructive pulmonary disease patients. *International Journal of COPD* 2015; 10: 745-757.

[54] Henoch, I, Strang, S, Löfdahl, C *et al*. Health-related quality of life in a nationwide cohort of patients with COPD related to other characteristics. *European Clinical Respiratory Journal* 2016; 3: 31459.

[55] Yoo, JY, Kim, YS, Kim, SS *et al*. Factors affecting the trajectory of health-related quality of life in COPD patients. *International Journal of Tuberculosis and Lung Disease* 2015; 20(6): 738-746.

[56] Jones, P, Lareau, S and Mahler, DA. Measuring the effects of COPD on the patient. *Respiratory Medicine* 2005; 99(Suppl B): S11-S18.

[57] Rubinsztajn, R, Przybylowski, T, Maskey-Warzechowska, M *et al*. Exacerbations of Chronic Obstructive Pulmonary Disease and Quality of Life of Patients. *Advances in Experimental Medicine, Biology - Neuroscience and Respiration* 2016; 884: 69-74.

[58] Kim, JK, Lee, SH, Lee, BH *et al*. Factors associated with exacerbation in mild-to-moderate COPD patients. *International Journal of COPD* 2016; 11: 1327-1333.

[59] Chandra, D, Tsai, CL and Camargo Jr, CA. Acute Exacerbations of COPD: Delay in Presentation and the Risk of Hospitalization. *COPD: Journal of Chronic Obstructive Pulmonary Disease* 2009; 6(2): 95-103.

[60] Van den Akker, EFMM, van 't Hul, A, Birnie, E *et al*. Comprehensive Diagnostic Assessment of Health Status of Patients with Asthma or COPD: A Delphi Panel Study among Dutch Experts. *COPD: Journal of Chronic Obstructive Pulmonary Disease* 2017; 14(2): 190-199.

[61] Global initiative for Obstructive Lung Disease. Pocket Guide to COPD Diagnosis, Management and Prevention: A Guide for Health Care Professional 2017. Accessed 11th of July 2018 from https://goldcopd.org/wp-content/uploads/2016/12/wms-GOLD-2017-Pocket-Guide-1.pdf.

[62] Jones, PW, Nadeau, G, Small, M *et al*. Characteristics of a COPD population categorized using the GOLD framework by health status and exacerbations. *Respiratory Medicine* 2014; 108(1): 129-135.

[63] Jones, PW, Adamek, L, Nadeau, G *et al*. Comparisons of health status scores with MRC grades in COPD: implications for the GOLD 2011 classification. *European Respiratory Journal* 2013; 42(3): 647-654.

[64] Han, MK, Muellerova, H, Curran-Everett, D *et al*. GOLD 2011 disease severity classification in COPDGene: a prospective cohort study. *Lancet Respiratory Medicine* 2013; 1(1): 43-50.

[65] Huang, WC, Wu, MF, Chen, HC *et al*. Features of COPD patients by comparing CAT with mMRC: a retrospective, cross-sectional study. *npj Primary Care Respiratory Medicine* 2015; 25: 15063.

[66] Karloh, M, Mayer, AF, Maurici, R *et al*. The COPD Assessment Test: what do we know so far? A systematic review and meta-analysis about clinical outcomes prediction and classification of patients into GOLD stages. *Chest* 2016; 149(2): 413-425.

[67] Casanova, C, Marin, JM, Martinez-Gonzalez, C *et al*. Differential Effect of Modified Medical Research Council Dyspnea, COPD Assessment Test, and Clinical COPD Questionnaire for Symptoms Evaluation Within the New GOLD Staging and Mortality in COPD. *Chest* 2015; 148(1): 159-168.

[68] Tsiligianni, IG, Alma, HJ, De Jong, C *et al*. Investigating sensitivity, specificity, and area under the curve of the Clinical COPD Questionnaire, COPD Assessment Test, and Modified Medical Research Council scale according to GOLD using St George's Respiratory Questionnaire cutoff 25 (and 20) as reference. *International Journal of COPD* 2016; 11: 1045-1052.

[69] Smid, DE, Spruit, MA, Deeg, DJH *et al*. How to determine an impaired health status in COPD: Results from a population-based study. *Netherlands Journal of Medicine* 2017; 75(4): 151-157.

[70] Smid DE, Franssen, FME, Gonik, M *et al*. Redefining Cut-Points for High Symptom Burden of the Global Initiative for Chronic Obstructive Lung Disease Classification in 18,577 Patients With Chronic Obstructive Pulmonary Disease. *Journal of the American Medical Directors Association (JAMDA)* 2017; 18(12): 1097.

[71] Jones, PW, Price, D and van der Molen, T. Role of clinical questionnaires in optimizing everyday care of chronic obstructive pulmonary disease. *International Journal of COPD* 2011; 6: 289-296.

[72] Cave, J, Atkinson, L, Tsiligianni, IG *et al.* Assessment of COPD wellness tools for use in primary care: an IPCRG initiative. *International Journal of COPD* 2012; 7: 447-456.

[73] Jones, PW, Harding, G, Berry, P *et al*. Development and first validation of the COPD Assessment Test. *European Respiratory Journal* 2009; 34(3): 648-654.

[74] Jones, PW, Bruselle, G, Dal Negro, RW *et al*. Properties of the COPD assessment test in a cross-sectional European study. *European Respiratory Journal* 2011; 38(1): 29-35.

[75] Dodd, JW, Hogg, L, Nolan, J *et al.* The COPD assessment test (CAT): response to pulmonary rehabilitation. A multicenter, prospective study. *Thorax* 2011; 66(5): 425-429.

[76] Dodd, JW, Marns, PL, Clark, AL *et al*. The COPD Assessment Test (CAT): Short- and Medium-term Response to Pulmonary Rehabilitation. *COPD: Journal of Chronic Obstructive Pulmonary Disease* 2012; 9(4): 390-394.

[77] Kon, SSC, Clark, AL, Dilaver, D *et al*. Response of the COPD Assessment Test to pulmonary rehabilitation in unselected chronic respiratory disease. *Respirology* 2013; 18(6): 974-977.

[78] Candemir, I, Kaymaz, D, Ergun, P *et al*. Assessment of pulmonary rehabilitation efficacy in chronic obstructive pulmonary disease patients using the chronic obstructive pulmonary disease assessment test. *Expert Review Respiratory Medicine* 2015; 9(4): 487-492.

[79] Chaplin, E, Gibb, M, Sewell, L *et al*. Response of the COPD Assessment Tool in Stable and Post-exacerbation Pulmonary Rehabilitation Populations. *Journal of Cardiopulmonary Rehabilitation and Prevention* 2015; 35(3): 214-218.

[80] Jones PW, Harding, G, Wiklung, I *et al*. Tests of the Responsiveness of the COPD Assessment Test Following Acute Exacerbation and Pulmonary Rehabilitation. *Chest* 2012; 142(1): 134-140.

[81] Papaioannou, M, Pitsiou, G, Manika, K *et al*. COPD Assessment Test: A Simple Tool to Evaluate Disease Severity and Response to Treatment. *COPD* 2014; 11(5): 489-495.

[82] Lopez-Campos, JL, Fernandez-Villar, A, Calero-Acuna, C *et al*. Evaluation of the COPD Assessment Test and GOLD patient types: a cross-sectional analysis. *International Journal of COPD* 2015; 10: 975-984.

[83] Gupta, N, Pinto, L, Benedetti, A *et al.* The COPD Assessment Test: can it discriminate across COPD subpopulations? *Chest* 2016; 150(5): 1069-1079.

[84] Gupta, N, Pinto, LM, Morogan, A *et al*. The COPD assessment test: a systematic review. *European Respiratory Journal* 2014; 44(4): 873-884.

[85] Van der Molen, T, Willemse, BWM, Schokker, S *et al*. Development, validity and responsiveness of the Clinical COPD Questionnaire. *Health and Quality of Life Outcomes* 2003; 1: 13.

[86] Reda, AA, Kotz, D, Kocks, JWH *et al*. Reliability and validity of the clinical COPD questionnaire and chronic respiratory questionnaire. *Respiratory Medicine* 2010; 104(11): 1675-1682.

[87] Ställberg, B, Nokela, M, Ehrs, P *et al*. Validation of the Clinical COPD Questionnaire (CCQ) in primary care. *Health and Quality of Life Outcomes* 2009, 7: 26.

[88] Kocks, JWH, Kerstjens, HAM, Snijders, SL *et al.* Health status in routine clinical practice: validity of the clinical COPD questionnaire at the individual patient level. *Health and Quality of Life Outcomes* 2010; 8: 135.

[89] Zhou, Z, Zhou, A, Zhao, Y *et al*. Evaluating the Clinical COPD Questionnaire: A systematic review. *Respirology* 2017; 22(2): 251-262.

[90] Tsiligianni, IG, van der Molen, T, Moraitaki, D *et al*. Assessing health status in COPD. A head-to-head comparison between the COPD assessment test (CAT) and the clinical COPD questionnaire (CCQ). *BMC Pulmonary Medicine* 2012, 12: 20.

[91] Sundh, J, Ställberg, B, Lisspers, K *et al.* Comparison of the COPD Assessment Test (CAT) and the Clinical COPD Questionnaire (CCQ) in a Clinical Population. *COPD* 2016; 13(1): 57-65.

[92] Ringbaek, T, Martinez, G and Lange, P. A Comparison of the Assessment of Quality of Life with CAT, CCQ, and SGRQ in COPD Patients Participating in Pulmonary Rehabilitation. *COPD* 2012; 9(1): 12-15.

[93] Jones, PW, Quirck, FH and Baveystock, CM. The St George's Respiratory Questionnaire. *Respiratory Medicine* 1991; 85 (Suppl B): 25-31.

[94] Jones, PW, Quirck, FH, Baveystock, M *et al*. A Self-complete Measure of Health Status for Chronic Airflow Limitation: The St. George's Respiratory Questionnaire. *American Review of Respiratory Disease* 1992; 145(6): 1321-1327.

[95] Meguro, M, Barley, EA, Spencer, S *et al*. Development and Validation of an Improved, COPD-Specific Version of the St. George Respiratory Questionnaire. *Chest* 2007; 132(2): 456-463.

[96] Barr, JT, Schumacher, GE, Freeman, S *et al*. American Translation, Modification, and Validation of the St. George's Respiratory Questionnaire. *Clinical Therapeutics* 2000; 22(9): 1121-1145.

[97] Rutten van Mölken, M, Roos, B and Van Noord, JA. An empirical comparison of the St George's Respiratory Questionnaire (SGRQ) and the Chronic Respiratory Disease Questionnaire (CRQ) in clinical trial setting 1999. *Thorax* 1999; 54(11): 995-1003.

[98] Puhan, MA, Guyatt, GH, Goldstein, R *et al*. Relative responsiveness of the Chronic Respiratory Questionnaire, St. Georges Respiratory Questionnaire and four other health-related quality of life instruments for patients with chronic lung disease. *Respiratory Medicine* 2007; 101(2): 308-316.

[99] Jones, PW. Activity Limitation and Quality of life In COPD. *COPD: Journal of Chronic Obstructive Pulmonary Disease* 2007; 4(3): 273-278.

[100] Von Leupoldt, A, Reijnders, T, Schuler, M *et al*. Validity of a Self-administered Questionnaire Version of the Transition Dyspnea Index in Patients with COPD. *COPD* 2017; 14(1): 66-71.

[101] Mahler, DA and Witek, TJ. The MCID of the Transition Dyspnea Index is a Total Score of One Unit. *COPD: Journal of Chronic Obstructive Pulmonary Disease* 2005; 2(1): 99-103.

[102] Witek Jr, TJ and Mahler, DA. Minimal important difference of the transition dyspnoea index in a multinational clinical trial. *European Respiratory Journal* 2003; 21(2): 267-272.

[103] Cazzola, M, Hanania, NA, MacNee, W *et al*. A review of the most common patient-reported outcomes in COPD – revisiting current knowledge and estimating future challenges. *International Journal of COPD* 2015; 10: 725-738.

[104] Haughney, J and Gruffydd-Jones, K. Patient-centred outcomes in primary care management of COPD - what do recent clinical trial data tell us? *Primary Care Respiratory Journal* 2004; 13(4): 185-197.

[105] Weldam, SWM, Schuurmans, MJ, Liu, R *et al*. Evaluation of Quality of Life Instruments for use in COPD care and research: A Systematic Review. *International Journal of Nursing Studies* 2013; 50(5): 688-707.

[106] Guyatt, GH, Berman, LB, Townsend, M *et al*. A measure of quality of life for clinical trials in chronic lung disease. *Thorax* 1987; 42(10): 773-778.

[107] Paap, MCS, Lenferink, LIM, Herzog, N *et al*. The COPD-SIB: a newly developed disease-specific item bank to measure health-related quality of life in patients with chronic obstructive pulmonary disease. *Health and Quality of Life Outcomes* 2016; 14: 97.

[108] Van Stel, HF, Maillé, AR, Colland, VT *et al*. Interpretation of change and longitudinal validity of the quality of life for respiratory illness questionnaire (QoLRIQ) in inpatient pulmonary rehabilitation. *Quality of Life Research* 2003; 12(2): 133-145.

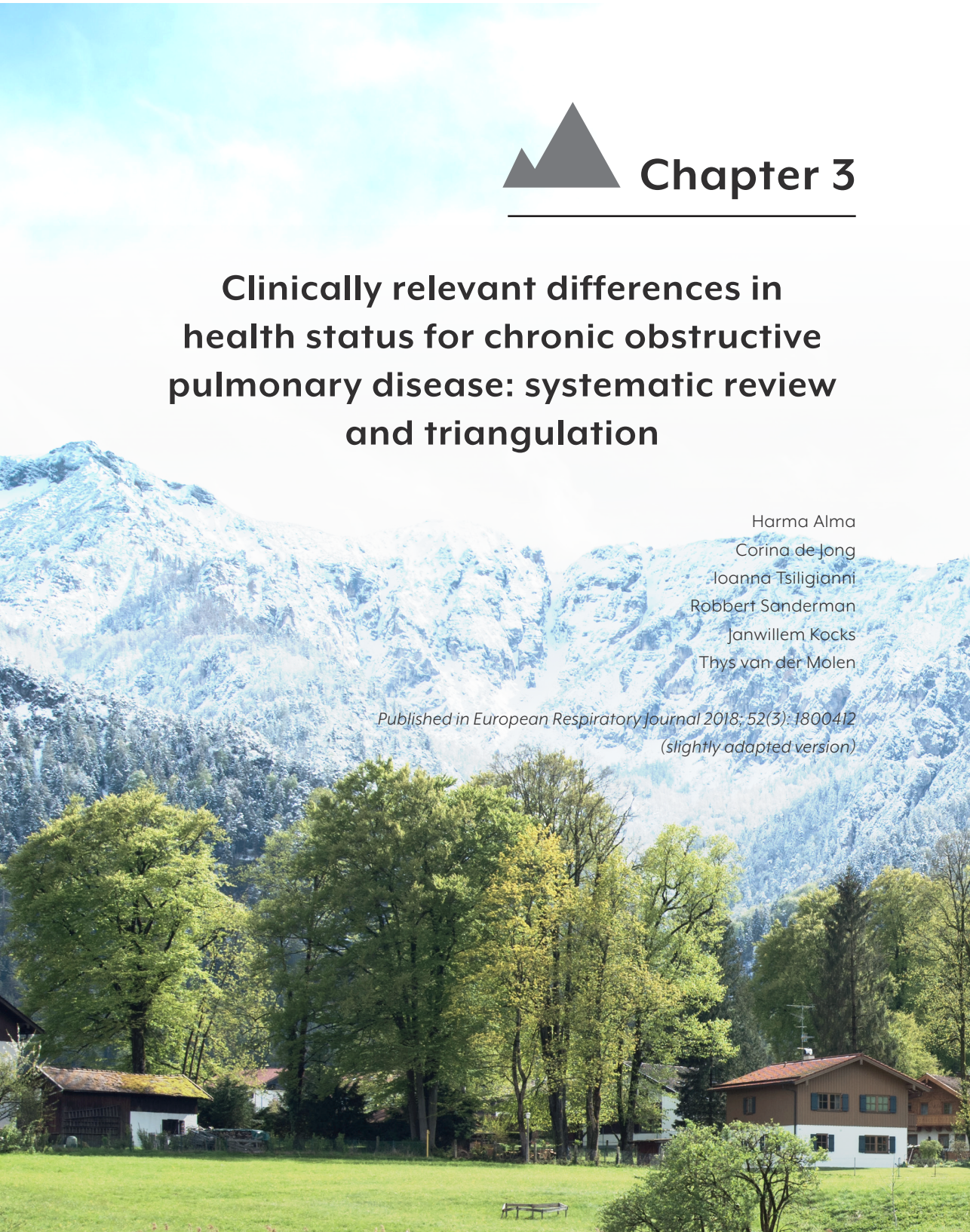
[109] Perez, T, Arnould, B, Grosbois, JM *et al*. Validity, reliability, and responsiveness of a new short Visual Simplified Respiratory Questionnaire (VSRQ) for health-related quality of life assessment in chronic obstructive pulmonary disease. *International Journal of Chronic Obstructive Pulmonary Disease* 2009; 4: 9-18.

[110] Mulhern, B, O'Gorman, H, Rotherham, N *et al*. Comparing the measurement equivalence of EQ-5D-5L across different modes of administration. *Health and Quality of Life Outcomes* 2015; 13: 191.

[111] Reardon, JZ, Lareau, SC and ZuWallack, R. Functional Status and Quality of Life in Chronic Obstructive Pulmonary Disease. *The American Journal of Medicine* 2006; 119(10 Suppl 1): S32–S37.

[112] Kocks, JWH, Asujee, GM, Tsiligianni, IG *et al*. Functional status measurement in COPD: a review of available methods and their feasibility in primary care. *Primary Care Respiratory Journal* 2011; 20(3): 269-275.

[113] Klijn, P, Legemaat, M, Beelen, A *et al*. Validity, Reliability, and Responsiveness of the Dutch Version of the London Chest Activity of Daily Living Scale in Patients With Severe COPD. *Medicine* 2015; 94(49):e2191.



Chapter 3

Clinically relevant differences in health status for chronic obstructive pulmonary disease: systematic review and triangulation

Harma Alma
Corina de Jong
Ioanna Tsiligianni
Robbert Sanderman
Janwillem Kocks
Thys van der Molen

Published in European Respiratory Journal 2018; 52(3): 1800412
(slightly adapted version)

3.1 Abstract

3.1.1 Background

The minimal clinically important difference (MCID) quantifies when measured differences can be considered clinically relevant. This study aims to review and triangulate MCIDs of health status tools for patients with chronic obstructive pulmonary disease (COPD).

3.1.2 Methods

A systematic search in PubMed, EMBASE and the Cochrane Library was conducted (PROSPERO #CRD42015023221). Study details, patient characteristics, MCID methodology and estimates were assessed and extracted by two authors. A triangulated mean was obtained for each tool's MCID, with two-thirds weighting for anchor-based and one-third for distribution-based results. This was then multiplied by a weighted factor based upon the study size and quality rating.

3.1.3 Results

Overall, 785 records were reviewed of which 21 studies were included for analysis. MCIDs of 12 tools were presented. General quality and risk of bias were average to good. Triangulated MCIDs for the COPD Assessment Test (CAT), Clinical COPD Questionnaire (CCQ) and St. George's Respiratory Questionnaire (SGRQ) were respectively -2.54, -0.43 and -7.43 for improvement. Too few and/or too diverse studies were present to triangulate MCIDs of the other tools.

3.1.4 Discussion and conclusions

Evidence for the MCID of the CAT and CCQ was strong and triangulation was valid. Currently used MCIDs in clinical practice for the SGRQ (4 points) and the Chronic Respiratory Questionnaire (0.5 point) did not match the reviewed content, for which the MCIDs were much higher. Using too low MCIDs may lead to an overestimation of the interpretation of treatment effects. MCIDs for deterioration were scarce, which highlights the need for more research.

3.2 Background

Health status measurements and thresholds for clinically important change are frequently used as obligatory endpoints in medical trials, scientific research and clinical practice to evaluate the effects of an intervention [1-6]. The minimal clinically important difference (MCID) is a pivotal parameter that quantifies this threshold for clinically relevant change. It has been defined by Jaeschke *et al.* [7] as "*the smallest difference in score in the domain of interest, which patients perceive as beneficial and which would mandate, in the absence of troublesome side-effects and excessive costs, a change in the patient's management*". The MCID is often used as a cutoff value to compare the percentage of patients achieving the level of clinically relevant change after intervention or pharmacotherapy in comparison to a control group [5-6, 8]. It is also used to define sample size and to evaluate change in clinical practice. Setting an MCID too high could lead to underestimation of the interpretation of treatment effects; defining a MCID too low may result in overestimation of this interpretation.

The measurement of health status and its MCIDs is of particular interest in chronic obstructive pulmonary disease (COPD). Physiologic measures like spirometry are often of interest to the physician, but of limited importance to patients, because these outcomes do not correlate well with their quality of life (QoL) [3, 9-16]. Patients can have similar spirometry or blood tests, but may experience very different levels of QoL and health status. QoL is "*the degree of satisfaction or dissatisfaction with various aspects of life that may be important to the individual*" [17]. Health status is considered *"the impact of health on a person's ability to perform and derive fulfillment from the activities of daily life"* [10, 18]. It is a standardised way of measuring the concepts health-related quality of life (HRQoL), functional status and mental well-being [2, 10]. HRQoL and health status questionnaires – often patient-reported outcomes (PROs) – have received much attention in the last few years, resulting in their inclusion in the global initiative for chronic obstructive lung disease (GOLD) guidelines for the classification of patient risk groups to guide treatment [12].

Many general and disease-specific HRQoL and health status tools exist, with varying designs ranging from single items to complex multi-domain questionnaires [1-2, 9-11, 19-22]. It is important that an instrument has strong measurement properties, including responsiveness, interpretability and good signal-to-noise ratio [1, 9-10, 22-23]. The MCID is an important parameter within these categories. Many authors have discussed the theory and methods to determine an instrument's MCID [2, 4, 24-45]. These are generally divided into anchor-, distribution-, and opinion-based approaches. Each method has its pros and cons. To date, there is no gold standard in defining an instrument's MCID [2, 44]. Hence, many different practices occur, some better than others. It is recommended

that both anchor- and distribution-based methods are used, combined with evidence from clinical trial data and qualitative approaches, with a systematic review or expert panel to aim for triangulation [33, 45].

Given the importance of MCIDs for research and clinical practice, there has been an increase in studies investigating the MCID of HRQoL and health status tools for COPD. Various authors have provided overviews of frequently used instruments for COPD, incidentally reporting the value or a short description of its MCID [3, 8, 10, 16, 28-29, 38, 46-52]. In addition, some publications have presented a description of the evidence for the MCID of specific COPD-related outcomes, including the St. George's Respiratory Questionnaire (SGRQ) [53], exacerbations [39], Transition Dyspnoea Index (TDI) [54], forced expiratory volume in one second (FEV_1) [55], Clinical COPD Questionnaire (CCQ) [56], and COPD Assessment Test (CAT) [57]. Most studies have not recently been updated, and none of these studies attempted to evaluate the quality of the MCID methodology or aim for triangulation. Outside the field of COPD, systematic reviews have emerged that summarise, quantify and make a quality assessment of the MCIDs of PROs and functional status tests [31, 58-64]. Our study here is the first to do so within the field of COPD. We aimed to systematically review the available evidence for the MCID of various HRQoL and health status tools used in COPD practice; to assess the quality of their methodology; and to attempt to triangulate the results as a kind of meta-analysis.

3.3 Methods

3.3.1 Search strategy

The preferred reporting items for systematic reviews and meta-analyses (PRISMA) statement served as a guideline for this systematic review [65]. The study protocol was prepared and published via PROSPERO (#CRD42015023221) [66]. PubMed, EMBASE and the Cochrane Library were searched to identify relevant original full-text articles on the measurement of MCIDs of HRQoL and health status tools for patients with COPD. The search strategy included various terms for the MCID determination of HRQoL questionnaires and/or health status measurement tools and/or PROs in adults with COPD *(Supplementary material 3.7.1)*. The search was conducted on the 9th of June 2015 and updated regularly with the final update on the 16th of June 2017. It included all studies and research designs prior to this.

3.3.2 Study criteria

Studies were considered eligible if they included approaches and original measurement data for the MCID of a generic or disease-specific HRQoL questionnaire and/or health status instrument and/or PRO used in adults with COPD. HRQoL and health status instruments were considered eligible when they captured more than one domain of the concepts physical, psychological and social functioning [1-2]. For patients with COPD, this would include concepts such as breathlessness, fatigue, cough, sputum production, physical functioning, social functioning, mental well-being and exacerbations [16]. The term health status will be used for future reference in this review. Only full-text studies containing original data were included. Conference abstracts, editorials and opinion articles were excluded. Reviews were initially included to explore the references. Non-English publications were translated if considered eligible.

3.3.3 Study selection

Titles and abstracts of the identified articles were screened by two authors (HA and CdJ) independently. The screening process included: (1) the study design and type was identified; (2) the measurement tool was identified; (3) a judgement was made whether the tool was a questionnaire or PRO, which measured health status according to the predefined inclusion definition; (4) the population was identified and screened for adults with COPD; (5) the aim of the study was identified, which needed to determine the instrument's MCID; (6) a description of the MCID methodology and final quantitative estimates should be available; (7) final judgement for eligibility was made. Independent results from both authors were compared. Where disagreement occurred, this was discussed and consensus was reached; or a third author (IT, TvdM or RS) was consulted. Full-text articles were retrieved for the selected studies and again checked according to the above stated seven steps. The reference lists of the selected articles were screened for additional titles. The abstracts of the additional titles were screened accordingly for meeting the pre-defined inclusion criteria.

3.3.4 Quality assessment and risk of bias

Eligible full text articles were assessed for their quality and risk of bias by two authors (HA and CdJ) independently. Disagreement between the authors was solved by consensus or involvement of a third author (IT, TvdM or RS). The authors composed a quality assessment and risk of bias tool by selecting 31 relevant items from various sources, because there was no specific tool available for evaluating studies that measure an instrument's MCID *(Supplementary material 3.7.2)*. Furthermore, various research designs were included, which made it difficult to use one specific checklist.

Items on study methodology and questionnaire design were selected from the Cochrane Risk of Bias tool [67] and the COSMIN checklist [68]. These items concerned the attrition and missing data procedures; selective outcome reporting; risk of funding and ownership bias; availability of at least two health status measurements; time interval of measurement stated; similar test conditions for both measurements; follow-up completed; validation and properties of the health status tool described; floor- and ceiling effects described; whether the MCID was calculated; and whether criterion/ anchors used were considered golden standard. Additional items were retrieved from the systematic reviews by Bohannon *et al.* [58-59]: clear inclusion/exclusion criteria; systematic enrollment of patients; missing data percentage less than 25%; more than one anchor used; and the use of receiver operating characteristics (ROC) curves with an area under the curve (AUC) of at least 0.70. The current authors added the following items based upon recommendations in the literature [2, 33, 45, 69]: adequate description of the anchor and its properties; anchor correlations at least 0.50; global rating of change (GRC) used with 11 or more scoring options; type of clinical criterion used; more than one distribution-based method used; MCID for more than one population measured; and whether the MCID was determined for improvement, deterioration or both.

The general scoring of the quality assessment and risk of bias included the answering options "yes", "no", "unclear" and "not applicable", as deducted from the COSMIN checklist [68]. "Not applicable" was selected for MCID related items that were not relevant for the corresponding study. Positive answers / low bias items were scored 2 points; unclear items were scored 1 point; and negative answers / high bias / not applicable items were scored 0 points. Individual items were scored and presented. An overall total score with a maximum of 62 could be obtained. Five categories were defined for the overall quality stratification, which was required for triangulation procedures. The overall risk of bias and quality assessment was given a star rating between one and five, calculated from the summed scores as follows: 0–12, one star; 13–25, two stars; 26–37, three stars; 38–49, four stars; and 50–62, five stars.

3.3.5 Data extraction, synthesis and analysis

Data were extracted using a standardised form including the general article properties; study properties; patient characteristics; health status measurements; and MCID properties (methodology, type of change, type of MCID, MCID estimates, and missing data procedures). Results from the full-text analysis were categorised per identified health status tool. Data were presented in tables and figures. A narrative synthesis of the MCID results, its methodology and its quality was prepared per instrument including forest plots. Primary outcome measures were the quality assessment of the MCIDs for health status tools in COPD, an overview of its MCID methods and estimates; as well

as triangulation of the MCIDs where multiple studies per instrument existed. Since no standard for triangulation exists, the authors determined the final triangulation as following. Triangulation was executed by first determining an anchor-based and distribution-based MCID per included study. The anchor-based result received a weight of two-thirds, and the distribution-based method received a weight of one-third. The results were multiplied by a weighted factor depending on the study size (n) and the quality star rating (one to five stars). An overall triangulated mean MCID was calculated per health status tool.

3.4 Results

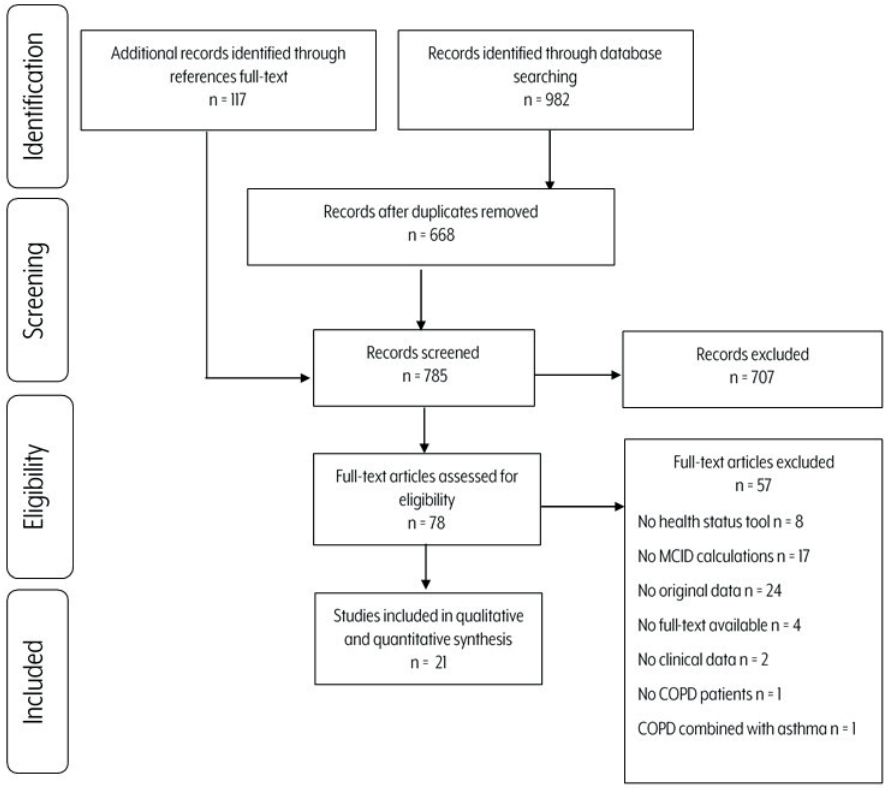
The initial search in PubMed, EMBASE and the Cochrane Library resulted in 668 unique studies (*Figure 1*). Screening the references provided an additional 117 titles, resulting in a total of 785 unique records. After screening all titles and abstracts, 78 papers were assessed for eligibility. A full-text analysis lead to the removal of 57 papers (*Online Supplementary Table 3, European Respiratory Journal*), leaving 21 records for inclusion (*Table 1*) [7, 70-89]. The initial level of agreement between the authors (HA and CdJ) was 89.2% for the study selection process. Cohens kappa for the quality and risk of bias assessment was 0.42. The included studies discussed in alphabetical order the tools CAT, CCQ, (short-form) Chronic Respiratory Questionnaire ((SF)-CRQ), the eDiary, EuroQol five Dimensions (EQ-5D), Feeling Thermometer (FT), SGRQ, Short-Form 6 Dimensions (SF-6D), Short-Form 36 (SF-36), Quality of Life for Respiratory Illness Questionnaire (QOLRIQ) and Visual Simplified Respiratory Questionnaire (VSRQ). Full patient characteristics, inclusion- and exclusion criteria, and health status scores are available in the *Online Supplementary Tables 4 and 5* (*European Respiratory Journal*).

Online Supplementary Tables

Most studies scored well on the overall quality of the general study methodology with the exception of the items on systematic enrollment, similar test conditions pre-and post-measurement, the description of floor- and ceiling effects, and unclear/high selective outcome reporting bias (*Table 2*). Four studies scored lower on the description of general study methodology [7, 72, 80, 87]. Regarding the quality of the MCID methodology, various studies scored poorly on the use of more than one anchor and/or lacked (sufficient) correlations with the chosen anchor [7, 71-78, 80, 85-88]. A limited number of studies used ROC curves in the anchor-based method of which most failed to define sufficient AUC [7,

71-75, 77-88]. A minority of studies used a GRC scale with sufficient answering categories [7, 70, 74, 79, 88]. Two studies used criterion referencing [70, 84]. Some of the included studies used more than one distribution-based method [70-71, 75-76, 82, 84-86]. A limited number of studies measured the MCID in more than one population [7, 72-73, 75, 87]. Certain studies determined the MCID for both improvement and deterioration [7, 71, 75, 80, 84-86, 88]. In most studies, there was a risk of ownership bias.

Figure 1: PRISMA flow-chart of the search results



Abbreviations: COPD, chronic obstructive pulmonary disease; MCID, minimal clinically important difference; N, number of items.

3.4.1 COPD Assessment Test (CAT)

The CAT contains eight questions with item scores ranging 0 (no limitations) up to 5 (maximum limitations) [90]. The total score derives from summing all items (min: 0, max: 40). Six papers discussed the MCID for the CAT in pulmonary rehabilitation (PR) [70, 72, 75, 82]; in patients with acute COPD exacerbation [73, 75]; and in regular primary and secondary care [75, 83] with follow-up periods of 2 weeks to 12 months (*Table 1*). The included studies received overall quality assessments of respectively two [83], three [72-73], four [75, 82]; or five stars [70] (*Table 2*). Anchor-based methods resulted in an MCID range for improvement for the CAT of -3.50 to -1.00 including the use of various GRCs by patients and physicians; exacerbations as a criterion; and CCQ, CRQ and SGRQ as anchors in ROC curves, linear regression analysis and mean change score calculations (*Figure 2 and Online Supplementary Table 6 European Respiratory Journal*) [70, 72, 73, 75, 82]. The anchor-based MCID for deterioration ranged +1 (n=51) to +2 (n=3) [72, 75]. Distribution-based approaches including the half standard deviation (0.5SD), standardised error of measurement (SEM) and 1.96 SEM ranged 1.92-3.80 [70, 75, 82, 83], excluding the 1.96SEM outlier of 6.43 [70]. The triangulated MCID for improvement was -2.54, excluding this 1.96SEM outlier. No structural differences were observed between different settings or follow-up periods. However, the anchor-based MCIDs by Dodd *et al.* [72] during 8 weeks of PR and by Kon *et al.* [75] during 12 months of regular care were smaller (*Figure 2*).

Online Supplementary Table 6

3.4.2 Clinical COPD Questionnaire (CCQ)

The CCQ contains 10 questions with item scores ranging 0 (no limitations) to 6 (maximum limitations) [91]. Total and domain scores (symptoms, functional and mental status) result from summing relevant scores and dividing this by the number of items (min: 0, max: 6). Five papers discussed the MCID for the CCQ in PR [70, 76, 82]; in patients with acute COPD exacerbation [74]; and in regular primary and secondary care [83] with follow-up periods of 2 days up to 12 months (*Table 1*). The overall quality of the included studies was rated as two [83], four [74, 76, 82], or five stars [70] (*Table 2*). The MCID for improvement for the CCQ total score ranged from -0.62 to -0.34 including various anchor-based methods with a 15-point GRC; criterion-referencing; and CAT, SGRQ and CRQ as anchors in linear regression, mean change calculations and ROC curves (*Figure 3, Online Supplementary Table 6 European Respiratory Journal*) [70, 74, 76, 82].

Table 1: Study characteristics of the included studies

First author	Year	N	Study setting	Study period	Follow-up period
COPD Assessment Test (CAT)					
Alma [70]	2016	449	Inpatient PR in Klinik Bad Reichenhall, Germany	February 2013 – July 2014	3 weeks
Dodd [72]	2011	297	Multidisciplinary PR in primary and secondary care in London, UK	January – August 2010	8 weeks
Jones [73]	2012	65	1) Regular care treatment for acute COPD exacerbation in primary and secondary care, USA	1) February 2009 – April 2009	2 weeks
		59	2) PR in Canada and USA	2) July 2009 – December 2009	6 weeks
Kon [75]	2014	565	1) Outpatient PR at Harefield Hospital, London, UK	1) April 2010 – December 2012	8 weeks
		147	2) Hospital discharge after admission from acute wards Hillingdon Hospital, London, UK	2) November 2011 – December 2012	3 months
		164	3) Regular care outpatient clinics Harefield Hospital, London, UK	3) January 2012 – August 2012	12 months
Smid [82]	2017	419	In- or outpatient PR at CIRO rehabilitation network Horn, the Netherlands	April 2012 – September 2014	8 or 16 weeks
Tsiligianni [83]	2012	90	Primary and secondary regular care Crete, Greece	July 2010 – June 2011	2 and 6 weeks
Clinical COPD Questionnaire (CCQ)					
Alma [70]	2016	449	Inpatient PR in Klinik Bad Reichenhall, Germany	February 2013 – July 2014	3 weeks
Kocks [74]	2008	168	Oral vs. i.v. prednisolone for acute COPD exacerbation in Isala klinieken, Zwolle, The Netherlands	June 2001 – May 2003	Days 1-7, 6 weeks, 12 months
Kon [76]	2014	261	Outpatient PR at Harefield Hospital, London, UK	November 2011 – January 2013	8 weeks
Smid [82]	2017	419	In- or outpatient PR at CIRO rehabilitation network Horn, the Netherlands	April 2012 – September 2014	8 or 16 weeks
Tsiligianni [83]	2012	90	Primary and secondary regular care Crete, Greece	July 2010 – June 2011	2 and 6 weeks
(Short-Form) Chronic Respiratory Questionnaire ((SF)CRQ)					
Chu-Lin Tsai [71]	2008	301	Regular care for acute COPD exacerbation in 29 EDs, USA and Canada	2000-2001	2 weeks
Jaeschke [7]	1989	28/23/20/10/21	1) Inpatient PR during 4-6 weeks	1) September 1983 – July 1984	Discharge and 2/6/12/24 weeks
		24	2) Clinical trial on inhaled salbutamol and oral theophylline for four 2-week periods vs. placebo	2) NR	4 x 2-week periods
Redelmeier [80]	1996	112	Small groups of patients in supervised PR	February 1992 – February 1994	Not applicable
Wyrwich [88]	2007	9	1) Expert panel discussion	1) 2000-2001	Every 2 months
		554/ 504/484/ 462/ 462	2) Regular primary care Wishard Health Services in Indianapolis, USA, and the St. Louis Veteran Affairs Medical Center	2) August 2000 – November 2001	Every 2 months
		43	3) Regular primary care Wishard Health Services in Indianapolis, USA, and the St. Louis Veteran Affairs Medical Center	3) August 2000 – November 2001	Not applicable

First author	Year	N	Study setting	Study period	Follow-up period
EQ-5D Utility Index (UI) and VAS					
Nolan [78]	2016	616	1) Regular care in respiratory clinics Harefield Hospital, London, UK	1) April 2012 – October 2014	Baseline
		324	2) Outpatient PR clinics Harefield Hospital, London, UK	2) August 2013 – October 2014	8 weeks
Walters [86]	2005	97/81	Regular care in chest clinic of a city teaching hospital	NR	6 / 12 months
Zanini [89]	2015	439	Inpatient PR tertiary healthcare center, Italy	January 2009 – December 2012	3 weeks
St. George's Respiratory Questionnaire (SGRQ)					
Alma [70]	2016	449	Inpatient PR in Klinik Bad Reichenhall, Germany	February 2013 – July 2014	3 weeks
Schünemann [81]	2003	84	PR patients at University of Toronto and McMaster University, Hamilton Ontario, Canada	NR	3 months
Tsiligianni [83]	2012	90	Primary and secondary regular care Crete, Greece	July 2010 – June 2011	2 and 6 weeks
Welling [87]	2015	110/86	Bronchoscopic lung volume reduction (BLVR) at University Medical Center Groningen, the Netherlands	NR	1 / 6 months
SF-6D and SF-36					
Walters [85]	2003	60	Regular care in chest clinic of a city teaching hospital	NR	1 year
Walters [86]	2005	97/81	Regular care in chest clinic of a city teaching hospital	NR	6 / 12 months
Wyrwich [88]	2007	9	1) Expert panel discussion	1) 2000-2001	Every 2 months
		554/ 504/484/ 462/ 462	2) Regular primary care Wishard Health Services in Indianapolis, USA, and the St. Louis Veteran Affairs Medical Center	2) August 2000 – November 2001	Every 2 months
		43	3) Regular primary care Wishard Health Services in Indianapolis, USA, and the St. Louis Veteran Affairs Medical Center	3) August 2000 – November 2001	Not applicable
Other tools:					
Kulich [77] – eDiary	2015	177	Phase III multi-center clinical trial (SHINE study) on QVA149 (dual bronchodilator)	September 2010 – February 2012	26 weeks
Perez [79] – VSRQ	2009	373	RCT 18 µg tiotropium once daily vs. placebo in 123 Centers in France	May 2002 – April 2004	2 weeks and 3 months
Schünemann [81] – FT	2003	84	PR patients at University of Toronto and McMaster University, Hamilton Ontario, Canada	NR	3 months
Van Stel [84] – QOLRIQ	2003	108	Individualised inpatient PR at Asthma Center Heideheuvel, the Netherlands	January 1996 – December 1997	3-6 months

Abbreviations: BLVR, bronchoscopic lung volume reduction; COPD, chronic obstructive pulmonary disease; ED, emergency department; N, number of patients; NR, not reported; PR, pulmonary rehabilitation; RCT, randomised controlled clinical trial; UK, United Kingdom; USA, United States of America.

Table 2: Quality assessment and risk of bias of the included studies

First author:	Inclusion criteria clear	Exclusion criteria clear	Systematic enrollment	Follow-up completed	Missing data reported	% Lost in follow-up <25%	Pre- and post measurements available	Time interval follow-up stated	Similar test conditions	Health status description	Validation health status tool	Floor effects	Ceiling effects	MCID calculation	Description anchor	Measurement properties anchor	Anchor golden standard	>1 Anchor used	Anchor correlations calculated	Anchor correlations ≥0.50	ROC curves produced	AUC ≥0.70	GRC used	Anchor question options ≥11	Minimal criterion used	>1 Distribution-based method	>1 Population used	Improvement & deterioration	Selective outcome reporting bias low	Funding bias risk low	Ownership bias risk low	Overall rating (1-5 stars)
(SF-)CRQ																															
Chu-Lin Tsai [71]	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●		●		●	●	●	●	●	●	●	●	★★★★
Jaeschke [7]	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●		●		●	●		●	●	●	●	●	★★★★
Redelmeier [80]	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●		●		●			●	●	●	●	●	★★
Wyrwich [88]	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●		●		●	●		●	●	●	●	●	★★★★
EQ-5D UI and VAS																															
Nolan [78]	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●		●	●	●	●	●	●	★★★★
Walters [86]	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●		●	●	●	●	●	●	●	●	★★★
Zanini [89]	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●		●	●	●	●	●	●	★★★★

First author:	Inclusion criteria clear	Exclusion criteria clear	Systematic enrollment	Follow-up completed	Missing data reported	% Lost in follow-up <25%	Pre- and post measurements available	Time interval follow-up stated	Similar test conditions	Health status description	Validation health status tool	Floor effects	Ceiling effects	MCID calculation	Description anchor	Measurement properties anchor	Anchor golden standard	>1 Anchor used	Anchor correlations calculated	Anchor correlations ≥0.50	ROC curves produced	AUC ≥0.70	GRC used	Anchor question options ≥11	Minimal criterion used	>1 Distribution-based method	>1 Population used	Improvement & deterioration	Selective outcome reporting bias low	Funding bias risk low	Ownership bias risk low	Overall rating (1-5 stars)
SGRQ																																
Alma [70]	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	★★★★★
Schünemann [81]	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●			●	●	●	●	●	★★★
Tsiligianni [83]	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●				●	●			●	●	●	●	●	★★
Welling [87]	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●			●	●	●	●	●	★★★
SF-6D and SF-36																																
Walters [85]	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	★★★
Walters [86]	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	★★★
Wyrwich [88]	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●		●	●	●	●		●	●	●	●	●	★★★★
Other tools																																
Kulich [77]	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	★★★
Perez [79]	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●			●	●	●	●	●	★★★★
Schünemann [81]	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●		●	●	●	●	●	●	★★★
Van Stel [84]	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	★★★★

Data overall: ● Respective item is clearly defined, poses a low risk of bias, or the considered test was performed. ● Respective item is not clearly defined, poses a high risk of bias, or this test was not performed. ● Respective item is unclear. ● Respective item is not clearly defined, poses a high risk of bias, or this test was not performed.

Data on the item "improvement & deterioration": ● ● both improvement and deterioration MCIDs. ● MCID only for improvement or deterioration calculated. ● Unclear whether MCID is for improvement, deterioration or both.

Abbreviations: AUC, area under the curve; CAT, COPD Assessment Test; CCQ, Clinical COPD Questionnaire; COPD, chronic obstructive pulmonary disease; CRQ, Chronic Respiratory Questionnaire; SF-CRQ, Short-Form Chronic Respiratory Questionnaire; EQ-5D, EuroQol 5 Dimensions; GRC, global rating of change; MCID, minimal clinically important difference; ROC, receiver operating characteristics; SGRQ, St. George's Respiratory Questionnaire; SF-6D, Short-Form 6 Dimensions; SF-36, Short-Form 36; UI, Utility Index; VAS, Visual Analogue Scale.

The MCID for deterioration for the CCQ has not been determined. Results on the domain scores are available in the *Online Supplementary Table 6* (*European Respiratory Journal)*. Distribution-based methods with 0.5SD, SEM and 1.96 SEM ranged from 0.21 to 0.56 [70, 74, 76, 82-83], excluding the outlier of the 0.80 estimate of the minimal detectable change 95% confidence interval (MDC95) [76]. The triangulated MCID for improvement for the CCQ was -0.43. Estimates from PR, patients with acute exacerbation and regular care with various follow-up durations were similar, except for the distribution-based estimate by Kocks *et al.* [74] in exacerbation patients (*Figure 3*).

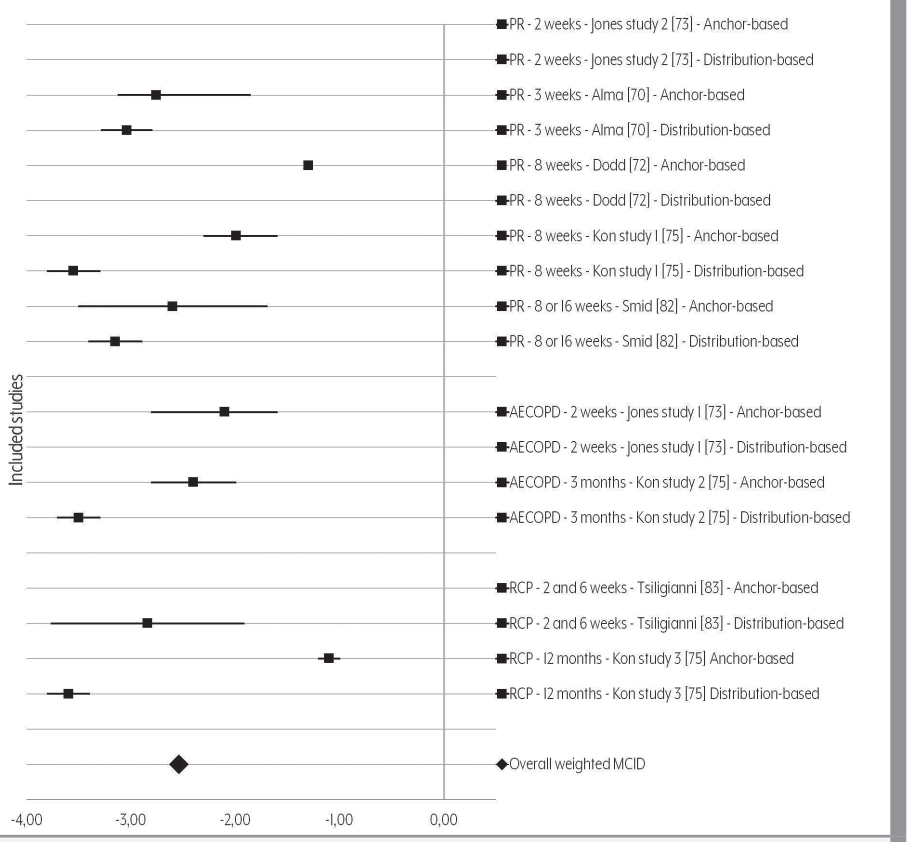
3.4.3 (SF-)CRQ

The CRQ consists of 20 items scored on a 7-point scale ranging 1 (most troubles) to 7 (no troubles) on the domains dyspnoea (5 items), fatigue (4 items), emotional function (7 items), and mastery (4 items) [92]. Domain scores are determined by summing the scores or determining the mean of the summed items [7, 80, 88]. The SF-CRQ includes 2 selected items per domain [71].

Four papers reported the MCID for the (SF)-CRQ in PR [7, 80]; in patients with acute COPD exacerbation [71]; in a salbutamol trial [7]; in regular primary care [88] and by means of expert opinions [88] (*Table 1*), with follow-up periods of 2–24 weeks. Overall quality was rated two [80], three [7, 71] or four stars [88] (*Table 2*). MCID estimates for both improvement and deterioration were determined in all studies (*Online Supplementary Table 6 European Respiratory Journal*).

The MCIDs resulted from the anchor-based method using a 5-, 7- or 15-point GRC for both within- and between subject change [71, 7, 88, 80]. The MCIDs for improvement for the SF-CRQ ranged from 0.30 to 1.60 as average domain score change (2 items per domain); and for deterioration from -0.60 to -0.06 [71]. The MCIDs for the CRQ for improvement per item score were: 0.40 to 1.00 (dyspnoea), 0.25 to 0.50 (fatigue), 0.14 (emotion), and 0 to 0.25 (mastery). The MCIDs for the CRQ for deterioration per item score were -0.20 (dyspnoea), -0.50 (fatigue), -0.14 to 0 (emotion) and -0.50 to -0.25 (mastery) [88]. A combined MCID for improvement and deterioration was per item 0.09 to 0.62 (dyspnoea), 0.50 to 0.68 (fatigue), 0.57 to 0.87 (emotion), and 0.23 to 0.75 (mastery) [7, 80, 88]. Due to the limited number of studies, the diversity of domains and scoring approaches, and the small number of patients in certain studies [7], no triangulation was performed.

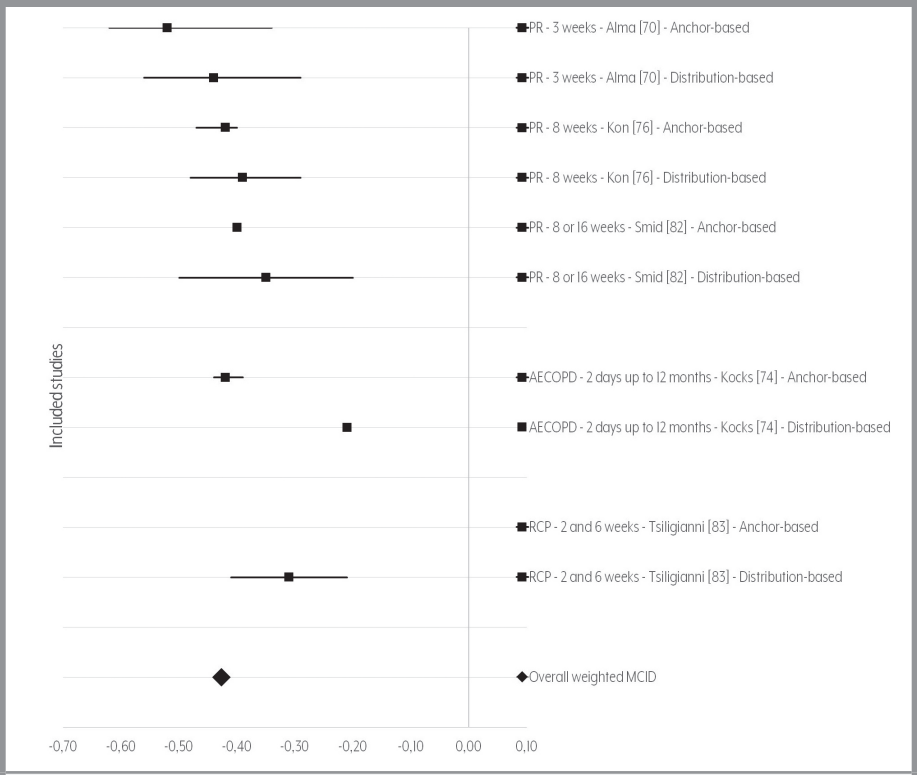
Figure 2: Overview study results and triangulation for the MCID of the CAT



Data presented as mean study MCIDs for anchor-based and distribution-based methods (squared estimates). The horizontal lines include the range of estimates provided in the respective study. The larger diamond represent the triangulated MCID. Results are categorised per setting in correspondence with the duration of follow-up period.

Abbreviations: AECOPD, acute exacerbation chronic obstructive pulmonary disease; CAT, COPD Assessment Test; COPD, chronic obstructive pulmonary disease; MCID, minimal clinically important difference; PR, pulmonary rehabilitation.

Figure 3: Overview study results and triangulation for the MCID of the CCQ



Data presented as mean study MCIDs for anchor-based and distribution-based methods (squared estimates). The horizontal lines include the range of estimates provided in the respective study. The larger diamond represent the triangulated MCID. Results are categorised per setting in correspondence with the duration of follow-up period.

Abbreviations: AECOPD, acute exacerbation chronic obstructive pulmonary disease; CCQ, Clinical COPD Questionnaire; COPD, chronic obstructive pulmonary disease; MCID, minimal clinically important difference; PR, pulmonary rehabilitation.

3.4.4 EQ-5D Utilities Index and Visual Analogue Scale

The EQ-5D contains the 5 dimensions mobility, self-care, usual activities, pain/discomfort, and anxiety/depression with each three (EQ-5D-3L) or five levels (EQ-5D-5L) in scoring severity [93-94]. A scoring algorithm results in an utility index (UI) between -0.590 (worst health) and +1.000 (best health) for the 3L version; and -0.208 to +1.000 for the 5L version. In addition, a visual analogue scale (VAS) score must be marked from 0 (worst health) to 100 (best health). Three papers discussed the MCID for the EQ-5D-5L UI, VAS and/or EQ-5D-3L-VAS in PR [78, 89]; and for the EQ-5D-3L in regular secondary care [86], with follow-up periods of 3 weeks to 12 months (*Table 1*). The quality assessment differed from three [86] to four stars [78, 89] (*Table 2*). The MCID for improvement for the EQ-5D-VAS ranged from 6.50 to 10.10 [78, 89]. The anchor-based MCID for improvement for the EQ-5D-3L

and EQ-5D-5L UI ranged from -0.128 to 0.063 [78, 86]. Estimates for deterioration ranged from -0.007 to 0.039 [86]. A combined MCID for improvement and deterioration ranged from -0.011 to zero [86]. Distribution-based results ranged from -0.050 to 0.150 [78, 86] (*Online Supplementary Table 6 European Respiratory Journal*). Owing to the limited number of included studies and the diversity of the results, no triangulation was executed.

3.4.5 SF-6D and SF-36

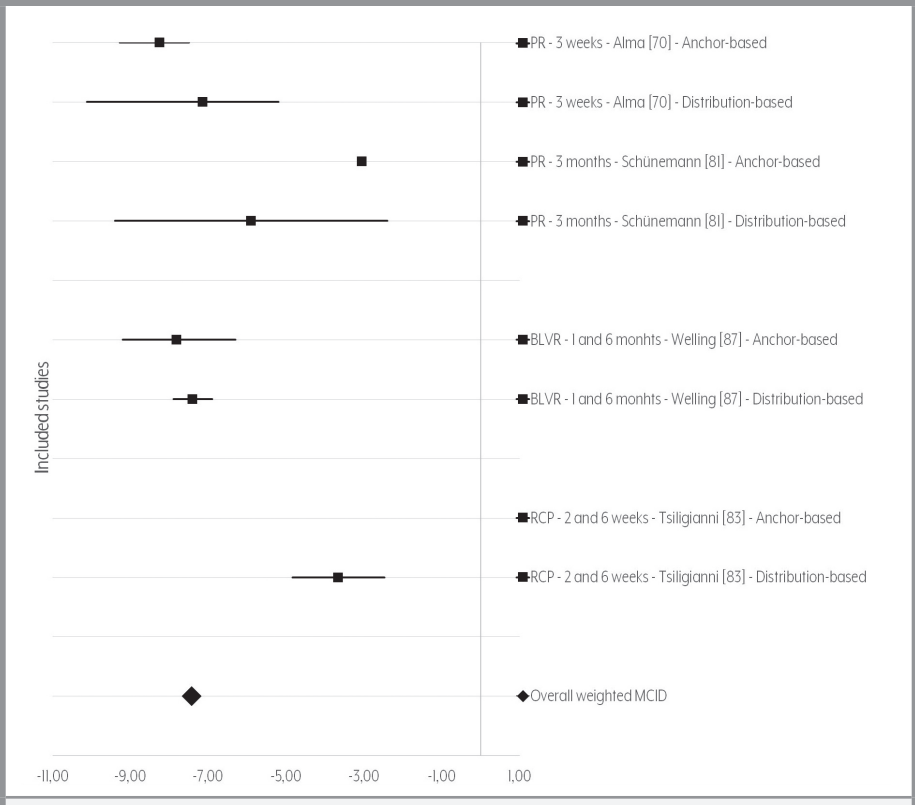
The SF-36 contains 36 items divided over eight domains each scoring between 0 (worst health) and 100 (best health) [95]. The SF-6D includes six dimensions resulting in a health state ranging 0.29 (worst health) to 1.00 (full health) [85]. Two papers discussed the MCID for the SF-6D in regular care [85-86], and one for the SF-36 in regular care and by means of expert opinions [88] (*Table 1*). Both studies on the SF-6D [85-86] were of average quality (three stars); the study on the SF-36 by Wyrwich *et al*. [88] was of good quality (four stars) (*Table 2*). The MCID for improvement for the various SF-36 domains ranged from 2 to 11 using a GRC scored by the patient or the physician [88]. The range for deterioration was -6 to +4. The expert-based panel determined that values of 8.33 to 12.50 to represent minimal changes [88]. The MCID for improvement for the SF-6D using a 5-point GRC ranged from -0.004 to +0.054; for deterioration ranged from 0.012 to 0.028; and combined ranged from 0.010 to 0.036 [85-86]. The distribution-based estimates for the SF-6D ranged from 0.044 to 0.410 using the standardised response mean (SRM), effect size (ES) and 0.5SD [85-86] (*Online Supplementary Table 6 European Respiratory Journal*). Owing to the limited number of included studies and diversity of the results, no triangulation was performed.

3.4.6 St. George's Respiratory Questionnaire (SGRQ)

The SGRQ is a 50-item questionnaire containing the domains symptoms, activities and impact with total and domain scores ranging 0 (best health status) to 100 (worst health status) [96]. Four studies analysed the MCID for the SGRQ during PR [70, 81]; bronchoscopic lung volume reduction (BLVR) [87]; and regular primary and secondary care [83] with follow-up periods of 2 weeks to 6 months (*Table 1*). The included studies scored two [83], three [81, 87], or five stars [70] (*Table 2*). The MCID was determined for improvement only. Anchor-based approaches resulted in an MCID for the SGRQ between -9.28 and -6.30 using a 15-point GRC, criterion-referencing; and CAT, CCQ, FEV_1, six minute walking distance (6MWD) and residual volume (RV) as anchors in linear regression, mean change calculations and ROC curves (*Figure 4, Online Supplementary Table 6 European Respiratory Journal*) [70, 87]. An outlying anchor-based result by Schünemann *et al.* [81] was -3.05 using the CRQ dyspnoea domain in linear regression analysis. Distribution-based results ranged from 2.40 to 10.19 using 0.2-0.8SD, SEM and 1.96SEM [70, 81, 83, 87]. The triangulated MCID for improvement was -7.43. Estimates from 3 weeks of PR [70],

and 1 and 6 months BLVR [87] were similar (*Figure 4*). However, the 3 months' PR anchor-based estimate by Schünemann *et al.* [81] was much smaller using the CRQ dyspnoea domain as anchor. The distribution-based result by Tsiligianni *et al.* [83] was also much smaller, when measured in regular care.

Figure 4: Overview study results and triangulation for the MCID of the SGRQ



Data presented as mean study MCIDs for anchor-based and distribution-based methods (squared estimates). The horizontal lines include the range of estimates provided in the respective study. The larger diamond represent the triangulated MCID. Results are categorised per setting in correspondence with the duration of follow-up period.

Abbreviations: BLVR, bronchoscopic lung volume reduction; MCID, minimal clinically important difference; PR, pulmonary rehabilitation, SGRQ, St. George's Respiratory Questionnaire.

3.4.7 Other tools

The other tools discussed included the eDiary [77], VSRQ [79], FT [81] and the QOLRIQ [84] (*Table 1*). The eDiary contains 5 symptom items and 2 impact items, resulting in scores ranging from 0 (best possible state/no problems) to 10 (worst possible state) [77]. The FT is a VAS ranging from 0 (worst state) to 100 (best score) [81]. The QOLRIQ contains 55 items regarding breathing problems, physical problems, emotions, situations triggering or enhancing breathing problems, general activities, daily and domestic activities, and social activities, relationships and sexuality [84]. Scores range on a 7-point scale with higher scores representing more impairment. The VSRQ contains 8 items covering dyspnoea, anxiety, depression, sleep, energy, daily activities, social activities, and sexual life [79]. Scores range from 0 to 10 with lower scores indicating higher impact on the patients' HRQoL. The included studies were of average (three stars) [77, 81] to good quality (four stars) [79, 84] (*Table 2*). All studies measured the MCID using anchor-based methods with multiple anchors using reasonable methodology. The MCID for improvement for the eDiary was -0.64 to -0.52 using a 7-point GRC and the TDI as anchors [77] (*Online Supplementary Table 6 European Respiratory Journal*). The MCID for improvement for the VSRQ was 3.20 to 3.50 using a 15-point GRC and the SGRQ as anchors in linear regression [79]. The MCID for the FT was 4.10 to 16.30 using the CRQ fatigue domain, and the SGRQ total and domain scores as anchors in linear regression; as well as the distribution-based methods 0.2-0.8SD [79]. The MCID for improvement for the QOLRIQ was 0.51 to 0.64; and for deterioration 0.37 to 0.49 using a 5-point GRC [84]. The distribution-based results ranged from 0.18 to 0.45 [84]. No triangulation was executed owing to the low number of studies per instrument.

3.5 Discussion

3.5.1 Summary of main results

The current systematic review provides a unique overview and triangulation of 21 papers including 12 different HRQoL and health status tools for COPD, and their MCID methodology, quality and estimates. The tools included are the CAT, CCQ, (SF)CRQ, eDiary, EQ-5D, FT, QOLRIQ, SF-6D, SF-36, SGRQ, and VSRQ. The overall quality of the methodology and MCID calculation was average to good, with one study scoring excellent and two studies scoring poor. Triangulated MCIDs for CAT, CCQ and SGRQ were -2.54, -0.43 and -7.43 for improvement, without structural differences between various settings and/or follow-up duration. The other instruments had too few or too heterogeneous studies to attempt triangulation; however ranges have been presented. Studies on MCIDs for deterioration were scarce or non-existent for all tools.

3.5.2 Interpretation of findings

COPD assessment and management should include health status instruments combined with the number of exacerbations to decide on patients' classification and therapy [12]. These tools are difficult to use in daily practice and scientific research without guidelines on what change may be considered clinically relevant [1, 3, 5, 9-10, 22-23]. The MCID parameter aims to quantify this threshold at the group-level [2, 6, 8]. It is an obligatory endpoint in clinical trials in which the percentage of patients achieving clinically relevant change is compared to the percentage of control patients achieving this change [2, 6, 8]. The MCID may also indicate to what extent an individual patient experiences relevant change over time. It is therefore of pivotal concern that the MCID is well established, else this may result in over- or underestimation of the interpretation of treatment effects. This review has provided insight into the quality and quantity of the MCID for various HRQoL and health status instruments for COPD with the CAT, CCQ, CRQ and SGRQ as the most important tools. The short CAT and CCQ are recommended especially for use in clinical practice; the lengthier CRQ and SGRQ are more applicable for scientific research [12, 52, 97].

The MCID for improvement for the CAT was between -4 and -1, with the majority of estimates between -3 and -2, resulting in a triangulated MCID of -2.54 [70, 72, 73, 75, 82-83]. This estimate was valid from multiple studies performed at the group level, demonstrating consistency of results for different settings with various follow-up periods. Stability of MCIDs during various follow-up periods was also demonstrated by Alma *et al.* [98]. Because the CAT only allows for integer scores at the individual level [90], a change of -3 could be considered a clinically relevant improvement for use in daily clinical practice. The MCID for deterioration for the CAT was between 1 and 2; however this resulted from two studies with a limited number of patients [72, 75]. The MCID for the CCQ was valid for improvement only, ranging from -0.60 to -0.20, leading to a triangulated estimate of -0.43 valid from PR, patients with an acute exacerbation, and regular care [70, 74, 76, 82-83]. Because the quality of the included studies for CAT and CCQ was average to excellent, and the quality and size of the study was integrated in the triangulation, these estimates are valid for use in clinical practice and scientific research. The triangulated estimates for CAT and CCQ are close to the accepted MCIDs currently used in practice nowadays, respectively -2 and -0.40 [12]. MCIDs for deterioration were not readily available.

The MCID for the SGRQ in the current review ranged from -11 to -2 for improvement, with most estimates between -10 and -6, resulting in a triangulated MCID of -7.43. However, the MCID for the SGRQ extensively used in clinical practice is 4 points. This estimate was based upon analyses of Jones and colleagues in 1991 and 2005 [53, 95]. The evidence

found in this review suggests the MCID to be almost double that and the studies that formed the basis for the currently accepted MCID for the SGRQ [53, 95] did not meet the study criteria for this review. It is therefore questionable how grounded the currently accepted MCID of 4 points is. Our evidence for the MCID for the SGRQ was of average to excellent quality, which validates the triangulated value of -7.43. The MCID for the SGRQ of 4 points for improvement has also been used as an anchor for the MCID of the CAT [70, 73, 75, 82], CCQ [70, 76, 82], VSRQ [79], and FT [81]. It may in fact have led to lower MCIDs for these tools. However, these tools have used other anchors and techniques in addition to the SGRQ to determine the MCID, validating their currently estimated (triangulated) values. Still, careful selection of anchors should be advocated.

CRQ item MCIDs of 0.50 points have been regularly used for both improvement and deterioration based upon Jaeschke *et al.* [7] and Redelmeier *et al.* [80]. However, the range of item MCIDs for both CRQ and SF-CRQ were wider and more inconsistent based upon the current review. The assessed quality of both studies was poor to average [7, 80]. The MCID for the CRQ of 0.5 might therefore be too simple an interpretation of results from methods of questionable quality. Owing to the inconsistent results, variety of scoring techniques and limited size of the studies, no triangulation was executed. The item MCID of 0.50 points has been used as an anchor for the MCIDs of the CAT [75], CCQ [76], SGRQ [81] and EQ-5D [78], which may have affected their MCIDs. The use of the CRQ as anchor did not result in structurally different results for the MCID of the CAT and CCQ. However, it did result in an outlying, possibly (too) low MCID for the SRGQ using the CRQ dyspnoea domain as reference [81].

The general health status instruments EQ-5D, SF-6D and SF-36 had less evidence for their MCIDs in COPD. The UI estimates for the EQ-5D and SF-6D varied and inconsistently ranged from minus to plus scores including the zero estimate. The MCID for improvement for the VAS of the EQ-5D was between 6 and 10 points [89]. General instruments may be applicable when comparing HRQoL between patient disease groups; however it is not valid to use their MCIDs to evaluate therapy outcomes within a patient group. To compare within patient groups, disease-specific health status tools such as CAT and CCQ are more valid, with well-established MCIDs. The other tools in our review, including the eDiary, FT, QOLRIQ, and VSRQ, each had only one study available regarding its MCID. More research is required for these instruments for them to be used in clinical practice.

In general, determining the MCID for an HRQoL and health status instrument requires a combination of anchor- and distribution-based methods, preferably measured in multiple settings over various follow-up periods [33, 45]. All included studies in this review, except for one [83], used anchor-based methods in measuring the MCID. Most studies,

except for five papers [7, 72-73, 80, 83], combined anchor-, distribution- and/or opinion-based methods as recommended. Studies regarding the CRQ, SF-6D and SF-36 in general did not use multiple anchors to determine the MCID. Furthermore, in most studies, the presentation of anchor correlations or correlations being ≥0.50, the use of ROC curves with AUC ≥0.70, and the use of a GRC with ≥11 answering options, was poor. Most studies did not use multiple distribution-based methods either. These would be points of attention for future MCID determination processes.

3.5.3 Strength and limitations of the current study

This study is the first to systematically address the MCIDs of HRQoL and health status tools for COPD. Although other papers have provided an overview of instruments, none of them has addressed the methodology, values and triangulation of MCIDs [3, 8, 10, 16, 28-29, 38-39, 46-57]. This study had a structurally defined protocol that was thoroughly executed by two independent reviewers. This review could be a starting point for further discussion. There are currently no fixed guidelines on how to judge studies measuring an instrument's MCID nor guidelines for triangulation. Triangulation should involve a combination of anchor- and distribution-based methods combined with evidence from clinical trial data and qualitative approaches [33, 45]; however this does not provide clear guidelines. The current authors have selected elements from existing assessment tools for their own risk of bias and quality assessment to evaluate the studies, its MCID methodologies and quality. This combined tool has not been validated. However, because it contained elements of established checklists and the MCID literature, the authors feel that there has been a sound evaluation of the overall quality and risk of bias of the included studies. The triangulation process as performed in the current review has not been used before, but takes into account the MCID methodology, its quality and the study size.

3.5.4 Implications for clinical practice and future research

This systematic review highlights gaps in the current MCID evidence and a need for further research. First, evidence is limited or lacking for the MCID for deterioration for all HRQoL and health status tools for COPD. This is relevant, because MCIDs for improvement are not necessarily similar to those for deterioration [2]. COPD is a progressive disease causing deterioration of HRQoL over time, which makes MCIDs for deterioration clinically important [99-100]. Second, the MCID for improvement for the CAT and CCQ were well-established in different settings for various follow-up durations. The triangulated MCIDs could be used in diverse clinical practices. However, the frequently used MCIDs for SGRQ and CRQ lacked evidence for the currently used clinical values of, respectively, 4 and 0.50 points. These estimates have been used as obligatory endpoints in clinical trials for many years. This might have resulted in an overestimation of the interpretation of

treatment effects. MCIDs for SGRQ and CRQ should thus be reconsidered; however this has major clinical consequences for the currently approved evidence-based therapies in COPD. It would be impossible to re-evaluate all clinical trials to date that have used the SGRQ and/or CRQ as outcome parameters. However, it may be worthwhile reviewing the existing evidence to observe which level of evidence remains valid with alternative MCIDs for the SGRQ and CRQ. Current knowledge and guidelines of health status measurement and its MCIDs have evolved over time. Older instruments such as the SGRQ and CRQ and their MCIDs may not have evolved with current guidelines. Last, evidence for the MCID for the eDiary, EQ5D, FT, QOLRIQ, SF-6D, SF-36, and VSRQ was limited in terms of consistency and number of studies available, highlighting the need for more research before they can reliably be used in clinical practice.

3.5.5 Conclusion

This study provides a first comprehensive and systematic assessment of MCIDs for HRQoL and health status instruments for patients with COPD. It highlights pros and cons in the used methodology, as well as gaps in the evidence. Triangulated MCIDs for the CAT, CCQ and SGRQ were, respectively, -2.54, -0.43 and -7.43 for improvement. These values may be integrated in future GOLD guidelines. This is an important step for clinicians and patients, who could easily use MCIDs in their scientific research and clinical practice. The MCIDs for the CAT and CCQ are well-established; however the reviewed MCIDs poorly matched with currently used values for the SGRQ (4 points) and CRQ (0.50 point). It is recommended that CAT or CCQ are used as outcome parameters for health status in COPD and that MCIDs for SGRQ and CRQ are recalculated. Evidence for the MCID for the other tools included was inconsistent, too heterogeneous or too limited. The fact that MCIDs for deterioration are scarce highlights a clear need for more well-designed studies.

3.6 Declarations

3.6.1 Funding

This review received financial support from the Junior Scientific Masterclass, part of the University of Groningen, to enable the PhD research position of Harma Alma.

3.6.2 Acknowledgements

We are grateful to the Junior Scientific Masterclass of the University of Groningen, who financially supported the research position of the first author. Furthermore, we wish to acknowledge mrs. D.G. van Ittersum at the University of Groningen for helping to define this review's search strategy.

3.6.3 Authors' contributions and consent

Harma Alma, Corina de Jong, Robbert Sanderman and Thys van der Molen designed the current systematic review. Harma Alma and Corina de Jong performed the search, article selection and data extraction procedures. Harma Alma wrote the first draft, while Corina de Jong, Ioanna Tsiligianni, Janwillem Kocks, Robbert Sanderman and Thys van der Molen actively participated in the review process. Robbert Sanderman and Thys van der Molen supervised and participated in different steps of the study, as well as in writing. All authors participated in various steps in the study, edited the manuscript and gave approval for submission.

3.6.4 Competing interests

Harma Alma, Corina de Jong and Robbert Sanderman have nothing to disclose. Janwillem Kocks reports personal fees from Novartis; research grants and personal fees from Boehringer Ingelheim; research grants and personal fees from GSK; research grants from Stichting Zorgdraad; personal fees from IPCRG; personal fees from Springer Media; and travel arrangements from Chiesi BV, GlaxoSmithKline BV, and IPCRG; all outside the submitted work. Ioanna Tsiligianni received personal fees from Boehringer Ingelheim, Novartis, AstraZeneca and GlaxoSmithKline; all outside the submitted work. Thys van der Molen reports personal reimbursements from GSK, TEVA, Astra Zeneca, Boehringer Ingelheim, and study grants from Astra Zeneca and GSK. This was all outside the submitted work. After this study was terminated, he became an employee of GSK. Thys van der Molen developed the CCQ and holds the copyright.

3.7 Supplementary material

3.7.1 Search strategy

Table 1: Search strategy

Database	Search terms	Search date
PubMed	P - Concept Patients with COPD "Pulmonary Disease, Chronic Obstructive"[Mesh] OR COPD [tw] OR Chronic Obstructive Pulmonary Dis* [tw] OR Obstructive Pulmonary Dis* [tw] OR Pulmonary Dis* [tw] OR Chronic Obstructive Airway Dis* [tw] OR Obstructive Airway Dis* [tw] OR Airflow Limitation* [tw] OR Airflow Obstruction* [tw] OR Chronic Bronchitis [tw] OR Bronchitis [tw] OR Emphysema [tw] OR Chronic Airway Dis* [tw] OR Respiratory Dis* [tw] AND I - Concept Patient reported health status questionnaires Patient-reported outcome*[tw] OR Patient Reported Outcome*[tw] OR PRO [tw] OR "Health Status"[Mesh] OR health status[tw] OR "Health Status Indicators"[Mesh] OR "Quality of Life"[Mesh] OR Quality Of Life [tw] OR QoL [tw] OR "Questionnaires"[Mesh] OR Questionnaire* [tw] AND C – none AND O – Minimal Clinically Important Difference (MCID) MCID [tw] OR MID [tw] OR minimum clinically important difference*[tw] OR minimum clinical important difference*[tw] OR minimum important difference*[tw] OR minimal clinically important difference*[tw] OR minimal clinical important difference*[tw] OR minimal important difference* [tw] OR minimally important difference*[tw] OR minimally clinical important difference*[tw] OR minimally clinically important difference*[tw] OR minimum clinically important change*[tw] OR minimum clinical important change* [tw] OR minimum important change*[tw] OR minimal clinically important change*[tw] OR minimal clinical important change* [tw] OR minimally important change*[tw] OR minimally clinically important change*[tw] OR minimally clinical important change* [tw] OR minimum clinically important improvement*[tw] OR minimum clinical important improvement* [tw] or minimum important improvement* [tw] OR minimal clinically important improvement*[tw] OR minimal clinical important improvement* [tw] or minimal important improvement* [tw] OR minimally clinical important improvement*[tw] OR minimally clinically important improvement*[tw]OR minimally important improvement*[tw] OR clinically meaningful difference*[tw] OR clinically meaningful change*[tw] OR clinically meaningful improvement*[tw] OR clinical meaningful difference*[tw] OR clinical meaningful change* [tw] OR clinical meaningful improvement* [tw]	Initial search on the 9th of June 2015. Updated on the 28th of January 2016 and the 13th of June 2017

Database	Search terms	Search date
EMBASE	P - Concept Patients with COPD 'obstructive airway disease'/exp OR 'chronic obstructive lung disease'/exp OR 'COPD':ab,ti OR 'Chronic Obstructive Pulmonary Disease':ab,ti OR 'Obstructive Pulmonary Disease':ab,ti OR 'Pulmonary Disease':ab,ti OR 'Chronic Obstructive Airway Disease':ab,ti OR 'Obstructive Airway Disease':ab,ti OR 'Airflow Limitation':ab,ti OR 'Airflow Obstruction':ab,ti OR 'Chronic Bronchitis':ab,ti OR 'Bronchitis':ab,ti OR 'Emphysema':ab,ti OR 'Chronic Airway Disease':ab,ti OR 'Respiratory Disease':ab,ti OR 'Chronic Obstructive Pulmonary Diseases':ab,ti OR 'Obstructive Pulmonary Diseases':ab,ti OR 'Pulmonary Diseases':ab,ti OR 'Chronic Obstructive Airway Diseases':ab,ti OR 'Obstructive Airway Diseases':ab,ti OR 'Airflow Limitations':ab,ti OR 'Airflow Obstructions':ab,ti OR 'Chronic Airway Diseases':ab,ti OR 'Respiratory Diseases':ab,ti OR 'Chronic Obstructive Pulmonary Disorder':ab,ti OR 'Obstructive Pulmonary Disorder':ab,ti OR 'Pulmonary Disorder':ab,ti OR 'Chronic Obstructive Airway Disorder':ab,ti OR 'Obstructive Airway Disorder':ab,ti OR 'Chronic Airway Disorder':ab,ti OR 'Respiratory Disorder':ab,ti OR 'Chronic Obstructive Pulmonary Disorders':ab,ti OR 'Obstructive Pulmonary Disorders':ab,ti OR 'Pulmonary Disorders':ab,ti OR 'Chronic Obstructive Airway Disorders':ab,ti OR 'Obstructive Airway Disorders':ab,ti OR 'Chronic Airway Disorders':ab,ti OR 'Respiratory Disorders':ab,ti AND I - Concept Patient reported health status questionnaires 'general health status assessment'/exp OR 'health status'/exp OR 'health status indicator'/exp OR 'quality of life'/exp OR 'quality of life assessment'/exp OR 'Patient Health Questionnaire'/exp OR 'questionnaire'/exp OR 'Patient-reported outcome':ab,ti OR 'Patient Reported Outcome':ab,ti OR 'PRO':ab,ti OR 'Health Status':ab,ti OR 'Quality Of Life':ab,ti or 'QoL':ab,ti OR 'Questionnaire':ab,ti OR 'Patient-reported outcomes':ab,ti OR 'Patient Reported Outcomes':ab,ti OR 'Questionnaires':ab,ti AND C - none AND O – Minimal Clinically Important Difference (MCID) 'MCID':ab,ti OR 'MID':ab,ti OR 'minimum clinically important difference':ab,ti OR 'minimum clinical important difference':ab,ti OR 'minimum important difference':ab,ti OR 'minimal clinically important difference':ab,ti OR 'minimal clinical important difference':ab,ti OR 'minimal important difference':ab,ti OR 'minimally important difference':ab,ti OR 'minimally clinical important difference':ab,ti OR 'minimally clinically important difference':ab,ti OR 'minimum clinically important change':ab,ti OR 'minimum clinical important change':ab,ti OR 'minimum important change':ab,ti OR 'minimal clinically important change':ab,ti OR 'minimal clinical important change':ab,ti OR 'minimally important change':ab,ti OR 'minimally clinically important change':ab,ti OR 'minimally clinical important change':ab,ti OR 'minimum clinically important improvement':ab,ti OR 'minimum clinical important improvement':ab,ti OR 'minimum important improvement':ab,ti OR 'minimal clinically important improvement':ab,ti OR 'minimal clinical important improvement':ab,ti OR 'minimal important improvement':ab,ti OR 'minimally clinical important improvement':ab,ti OR 'minimally clinically important improvement':ab,ti OR 'minimally important improvement':ab,ti OR 'clinically meaningful difference':ab,ti OR 'clinically meaningful change':ab,ti OR 'clinically meaningful improvement':ab,ti OR ->	Initial search on the 15th of June 2015. Updated on the 28th of January 2016 and the 13th of June 2017

Database	Search terms	Search date
	'clinical meaningful difference':ab,ti OR 'clinical meaningful change':ab,ti OR 'clinical meaningful improvement':ab,ti OR 'minimum clinically important differences':ab,ti OR 'minimum clinical important differences':ab,ti OR 'minimum important differences':ab,ti OR 'minimal clinically important differences':ab,ti OR 'minimal clinical important differences':ab,ti OR 'minimal important differences':ab,ti OR 'minimally important differences':ab,ti OR 'minimally clinical important differences':ab,ti OR 'minimally clinically important differences':ab,ti OR 'minimum clinically important changes':ab,ti OR 'minimum clinical important changes':ab,ti OR 'minimum important changes':ab,ti OR 'minimal clinically important changes':ab,ti OR 'minimal clinical important changes':ab,ti OR 'minimally important changes':ab,ti OR 'minimally clinically important changes':ab,ti OR 'minimally clinical important changes':ab,ti OR 'minimum clinically important improvements':ab,ti OR 'minimum clinical important improvements':ab,ti OR 'minimum important improvements':ab,ti OR 'minimal clinically important improvements':ab,ti OR 'minimal clinical important improvements':ab,ti OR 'minimal important improvements':ab,ti OR 'minimally clinical important improvements':ab,ti OR 'minimally clinically important improvements':ab,ti OR 'minimally important improvements':ab,ti OR 'clinically meaningful differences':ab,ti OR 'clinically meaningful changes':ab,ti OR 'clinically meaningful improvements':ab,ti OR 'clinical meaningful differences':ab,ti OR 'clinical meaningful changes':ab,ti OR 'clinical meaningful improvements':ab,ti AND [embase]/lim Conference abstracts were excluded using the filter option	

Database	Search terms	Search date
Cochrane Library	#1 MeSH descriptor: [Pulmonary Disease, Chronic Obstructive] explode all trees #2 COPD or "Chronic Obstructive Pulmonary Dis*" or "Obstructive Pulmonary Dis*" or "Pulmonary Dis*" or "Chronic Obstructive Airway Dis*" or "Obstructive Airway Dis*" or "Airflow Limitation*" or "Airflow Obstruction*" or "Chronic Bronchitis" or Bronchitis or Emphysema or "Chronic Airway Dis*" or "Respiratory Dis*":ti,ab,kw (Word variations have been searched) #3 MeSH descriptor: [Health Status] explode all trees #4 MeSH descriptor: [Health Status Indicators] explode all trees #5 MeSH descriptor: [Quality of Life] explode all trees #6 MeSH descriptor: [Questionnaires] explode all trees #7 "Patient-reported outcome*" or "Patient Reported Outcome*" or PRO or "health status" or "Quality of Life" or QoL or Questionnaire*:ti,ab,kw (Word variations have been searched) #8 MCID or MID or "minimum clinically important difference*" or "minimum clinical important difference*" or "minimum important difference*" or "minimal clinically important difference*" or "minimal clinical important difference*" or "minimal important difference*" or "minimally important difference*" or "minimally clinical important difference*" or "minimally clinically important difference*" or "minimum clinically important change*" or "minimum clinical important change*" or "minimum important change*" or "minimal clinically important change*" or "minimal clinical important change*" or "minimally important change*" or "minimally clinically important change*" or "minimally clinical important change*" or "minimum clinically important improvement*" or "minimum clinical important improvement*" or "minimum important improvement*" or "minimal clinically important improvement*" or "minimal clinical important improvement*" or "minimal important improvement*" or "minimally clinical important improvement*" or "minimally clinically important improvement*" or "minimally important improvement*" or "clinically meaningful difference*" or "clinically meaningful change*" or "clinically meaningful improvement*" or "clinical meaningful difference*" or "clinical meaningful change*" or "clinical meaningful improvement*":ti,ab,kw (Word variations have been searched) #9 (#1 OR #2) AND (#3 OR #4 OR #5 OR #6 OR #7) AND #8 Option trials was selected	Initial search on the 9th of July 2015. Updated on the 28th of January 2016 and the 16th of June 2017

3.7.2 Risk of bias and quality assessment

Table 2: Risk of bias and quality assessment form

Selected item	Scoring method (points)
1. Were participant inclusion criteria clearly defined?	Yes (2) No (0) Unclear (1)
2. Were participant exclusion criteria clearly defined?	Yes (2) No (0) Unclear (1)
3. Were patients systematically enrolled?	Yes (2) No (0) Unclear (1)
4. Was follow-up completed?	Yes (2) No (0) Unclear (1)
5. Were missing data procedures reported?	Yes (2) No (0) Unclear (1)
6. Which % lost in follow up?	<25% (2) ≥25% (0) Unclear (1)
7. Were at least two health status measurements (pre and post) available?	Yes (2) No (0) Unclear (1)
8. Was the time interval for follow-up stated?	Yes (2) No (0) Unclear (1)
9. Were test conditions similar for both measurements?	Yes (2) No (0) Unclear (1)
10. Was there an adequate description given of measurement instrument?	Yes (2) No (0) Unclear (1)
11. Was the instrument validated in the current study, or is made reference to other study?	Yes (2) No (0) Unclear (1)
12. Were floor effects described	Yes (2) No (0) Unclear (1)
13. Were ceiling effects described	Yes (2) No (0) Unclear (1)
14. Was the M(C)ID calculated?	Yes (2) No (0) Unclear (1)
15. Was an adequate description given of the anchor(s)?	Yes (2) No (0) Unclear (1) Not Applicable = N/A (0)
16. Were measurement properties of the anchor(s) described?	Yes (2) No (0) Unclear (1) Not Applicable = N/A (0)
17. Can the anchor(s) be considered a gold standard?	Yes (2) No (0) Unclear (1) Not Applicable = N/A (0)
18. Were >1 anchor used to determine M(C)ID?	Yes (2) No (0) Unclear (1) Not Applicable = N/A (0)
19. Were anchor correlations calculated?	Yes (2) No (0) Unclear (1) Not Applicable = N/A (0)
20. Were anchor correlations ≥0.50?	Yes (2) No (0) Unclear (1) Not Applicable = N/A (0)
21. Were receiver operating characteristics (ROC) curves produced?	Yes (2) No (0) Unclear (1) Not Applicable = N/A (0)
22. Was the area under the curve (AUC) ≥0.70?	Yes (2) No (0) Unclear (1) Not Applicable = N/A (0)
23. Was a global rating of change (GRC) used?	Yes (2) No (0) Unclear (1) Not Applicable = N/A (0)
24. Number of GRC anchor questions?	<11 (0) ≥11 (2) Unclear (1) Not Applicable = N/A (0)
25. What criterion was used?	Exacerbation (2) Hospital admission (1) Death (0) Other (1) Not Applicable = N/A (0)
26. Was more than one distribution-based method used?	Yes (2) No (0) Unclear (1) Not Applicable = N/A (0)
27. Was more than one population used in MCID?	Yes (2) No (0) Unclear (1)
28. MCID calculated for:	Improvement (1) Deterioration (1) Both (2)
29. Was there selective outcome reporting?	Yes (0) No (2) Unclear (1)
30. Was there funding bias?	Yes (0) No (2) Unclear (1)
31. Was there ownership bias?	Yes (0) No (2) Unclear (1)

3.8 References

[1] US Department of Health and Human Services FDA Center for Drug Evaluation and Research. Guidance for industry: patient-reported outcome measures: use in medical product development to support labelling claims: draft guidance. *Health and Quality of Life Outcomes* 2006; 4: 79.

[2] Crosby, RD, Kolotkin, RL and Williams, GR. Defining clinically meaningful change in health-related quality of life. *Journal of Clinical Epidemiology* 2003; 56(5): 395-407.

[3] Jones, PW. Health status measurement in chronic obstructive pulmonary disease. Thorax 2001; 56(11): 880-887.

[4] Beaton, DE, Boers, M and Wells, GA. Many faces of the minimal clinically important difference (MCID): a literature review and directions for future research. *Current Opinion in Rheumatology* 2002; 14(2): 109-114.

[5] Jones, PW. Health Status: What Does It Mean for Payers and Patients? *Proceedings of the American Thoracic Society* 2006; 3(3): 222-226.

[6] Jones, PW, Rennard, S, Tabberer, M *et al*. Interpreting patient-reported outcomes from clinical trials in COPD: a discussion. *International Journal of COPD* 2016; 11: 3069-3078.

[7] Jaeschke, R, Singer, J and Guyatt, GH. Measurement of health status. Ascertaining the Minimal Clinically Important Difference. *Controlled Clinical Trials* 1989; 10(4): 407-415.

[8] Jones, PW, Beeh, KM, Chapman, KR *et al*. Minimal Clinically Important Differences in Pharmacological Trials. *American Journal of Respiratory Critical Care Medicine* 2014; 189(3): 250-255.

[9] Guyatt, GH, Feeny, DH and Patrick, DL. Measuring health-related quality of life. *Annals of Internal Medicine* 1993; 118(8): 622-629.

[10] Curtis, JR and Patrick, DL. The assessment of health status among patients with COPD. *European Respiratory Journal Supplement* 2003; 41: 36s-45s.

[11] Jones, PW. Issues Concerning Health-Related Quality of Life in COPD. *Chest* 1995; 107(5 Suppl): 187S-193S.

[12] Global initiative for Chronic Obstructive Lung Disease (GOLD). Global Strategy for the Diagnosis, Management and Prevention of COPD (2017 Report). 2017. Available at http://www.goldcopd.org/gold-2017-global-strategy-diagnosis-management-prevention-copd/. Accessed 17th of November 2017.

[13] Westwood, M, Bourbeau, J, Jones, PW *et al*. Relationship between FEV_1 change and patient-reported outcomes in randomized trials of inhaled bronchodilators for stable COPD: a systematic review. *Respiratory Research* 2011; 12: 40.

[14] Tsiligianni, I, Kocks, J, Tzanakis, N *et al*. Factors that influence disease-specific quality of life or health status in patients with COPD: a review and meta-analysis of Pearson correlations. *Primary Care Respiratory Journal* 2011; 20(3); 257-268.

[15] Martin, AL, Marvel, J, Fahrbach, K *et al*. The association of lung function and St. George's respiratory questionnaire with exacerbations in COPD: a systematic literature review and regression analysis. *Respiratory Research* 2016; 17: 40.

[16] Jones, P, Miravittles, M, van der Molen, T *et al*. Beyond FEV_1 in COPD: a review of patient-reported outcomes and their measurement. *International Journal of COPD* 2012; 7: 697-709.

[17] Farquhar, M. Definitions of quality of life: a taxonomy. *Journal of Advanced Nursing* 1995; 22(3): 502-508.

[18] Reardon, JZ, Lareau, SC and ZuWallack, R. Functional Status and Quality of Life in Chronic Obstructive Pulmonary Disease. *The American Journal of Medicine* 2006; 119(10 Suppl 1): S32-37.

[19] Bergner, M and Rothman, ML. Health Status Measures: An Overview and Guide for Selection. *Annual Review of Public Health* 1987; 8: 191-210.

[20] Bergner, M. Quality of Life, Health Status and Clinical Research. *Medical Care* 1989; 27(3 Suppl): s148-156.

[21] Jones, P, Lareau, S and Mahler, DA. Measuring the effects of COPD on the patient. *Respiratory Medicine* 2005; 99(Suppl B): S11-S18.

[22] Fitzpatrick, R, Davey, C, Buxton, MJ *et al*. Evaluating patient-based outcome measures for use in clinical trials. *Health Technology Assessment* 1998; 2(14): 1-74.

[23] Davidson, M and Keating, J. Patient-reported outcome measures (PROMS): how should I interpret reports of measurement properties? A practical guide for clinicians and researchers who are not biostatisticians. *British Journal of Sports Medicine* 2014; 48(9): 792-796.

[24] Angst, F, Aeschlimann, A and Angst, J. The minimal clinically important difference raised the significance of outcome effects above the statistical level, with methodological implications for future studies. *Journal of Clinical Epidemiology* 2017; 82: 128-136.

[25] Copay, AG, Subach, BR, Glassman, SD *et al.* Understanding the minimum clinically important difference: a review of concepts and methods. *The Spine Journal* 2007; 7(5): 541-546.

[26] Hays, RD, Farivar, SS and Liu, H. Approaches and Recommendations for Estimating Minimally Important Differences for Health-Related Quality of Life Measures. *COPD: Journal of Chronic Obstructive Pulmonary Disease* 2005; 2(1): 63-67.

[27] Dworkin, RH, Turk, DC, Wyrwich, KW *et al.* Consensus Statement Interpreting the Clinical Importance of Treatment Outcomes in Chronic Pain Clinical Trials: IMMPACT Recommendations. *The Journal of Pain* 2008; 9(2): 105-121.

[28] Jones, PW. Interpreting thresholds for a clinically significant change in health status in asthma and COPD. *European Respiratory Journal* 2002; 19(3): 398-404.

[29] Make, B. How Can We Assess Outcomes of Clinical Trials: The MCID Approach. *COPD: Journal of Chronic Obstructive Pulmonary Disease* 2007; 4(3): 191-194.

[30] Norman, GR. The Relation Between the Minimally Important Difference and Patient Benefit. *COPD: Journal of Chronic Obstructive Pulmonary Disease* 2005; 2(1): 69-73.

[31] Ostelo, RWJG, Deyo, RA, Stratford, P *et al.* Interpreting Change Scores for Pain and Functional Status in Low Back Pain. Towards International Consensus Regarding Minimal Important Change. *Spine* 2008; 33(1): 90-94.

[32] Revicki, DA, Cella, D, Hays, RD *et al.* Responsiveness and minimal important differences for patient reported outcomes. *Health and Quality of Life Outcomes* 2006; 4: 70.

[33] Revicki, D, Hays, RD, Cella, D *et al.* Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *Journal of Clinical Epidemiology* 2008; 61(2): 102-109.

[34] Sloan, JA. Assessing the Minimally Clinically Significant Difference: Scientific Considerations, Challenges and Solutions. *COPD: Journal of Chronic Obstructive Pulmonary Disease* 2005; 2(1): 57-62.

[35] Turner, D, Schünemann, HJ, Griffith, LE *et al.* The minimal detectable change cannot reliably replace the minimal important difference. *Journal of Clinical Epidemiology* 2010; 63(1): 28-36.

[36] Wyrwich, KW and Wolinsky, FD. Identifying meaningful intra-individual change standards for health-related quality of life measures. *Journal of Evaluation in Clinical Practice* 2000; 6(1): 39-49.

[37] Wells, G, Beaton, D, Shea, B *et al.* Minimal Clinically Important Differences: Review of Methods. *The Journal of Rheumatology* 2001; 28(2): 406-412.

[38] Make, B, Casaburi, R and Leidy, NK. Interpreting Results from Clinical Trials: Understanding Minimal Clinically Important Differences in COPD Outcomes. *COPD: Journal of Chronic Obstructive Pulmonary Disease* 2005; 2(1): 1-5.

[39] Chapman, KR, Bergeron, C, Bhutani, M *et al.* Do We Know the Minimal Clinically Important Difference (MCID) for COPD Exacerbations? *COPD: Journal of Chronic Obstructive Pulmonary Disease* 2013; 10(2): 243-249.

[40] Hays, RD and Woolley, JM. The Concept of Clinically Meaningful Difference in Health-Related Quality of Life Research. How Meaningful is it? *Pharmacoeconomics* 2000; 18(5): 419-423.

[41] Middel, B. Statistical significant change versus relevant or important change in (quasi) experimental design: Some conceptual and methodological problems in estimating magnitude of intervention-related change in health services research. *International Journal of Integrated Care* 2002; 2: e15.

[42] Norman, GR, Sloan, JA and Wyrwich, KW. Interpretation of Changes in Health-related Quality of Life. The Remarkable Universality of Half a Standard Deviation. *Medical Care* 2003; 41(5): 582-592.

[43] Jacobson, NS and Truax, P. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology* 1991; 59(1): 12-19.

[44] Cook, CE. Clinimetrics Corner: The Minimal Clinically Important Change Score (MCID): A Necessary Pretense. *Journal of Manual & Manipulative Therapy* 2008; 16(4): 82-83.

[45] Leidy, NK and Wyrwich, KW. Bridging the Gap: Using Triangulation Methodology to Estimate Minimal Clinically Important Differences (MCIDs). *COPD: Journal of Chronic Obstructive Pulmonary Disease* 2005; 2(1): 157-165.

[46] Cazzola, M, Hanania, NA, MacNee, W *et al.* A review of the most common patient-reported outcomes in COPD – revisiting current knowledge and estimating future challenges. *International Journal of COPD* 2015; 10: 725-738.

[47] Haughney, J and Gruffydd-Jones, K. Patient-centred outcomes in primary care management of COPD - what do recent clinical trial data tell us? *Primary Care Respiratory Journal* 2004; 13(4): 185-197.

[48] Jayadevappa, R, Cook, R and Chatre, S. Minimal important difference to infer changes in health-related quality of life-a systematic review. *Journal of Clinical Epidemiology* 2017; 89: 188-198.

[49] Kocks, JWH, Asujee, GM, Tsiligianni, IG *et al*. Functional status measurement in COPD: a review of available methods and their feasibility in primary care. *Primary Care Respiratory Journal* 2011; 20(3): 269-275.

[50] Ekström, M, Currow, DC and Johnson, MJ. Outcome measurement of refractory breathlessness: endpoints and important differences. *Current Opinion in Supportive and Palliative Care* 2015; 9(3): 238-243.

[51] Weldam, SWM, Schuurmans, MJ, Liu, R *et al*. Evaluation of Quality of Life Instruments for use in COPD care and research: A Systematic Review. *International Journal of Nursing Studies* 2013; 50(5): 688-707.

[52] Kocks, JWH, Seys, SF, van Duin, TS *et al*. Assessing patient-reported outcomes in asthma and COPD patients: which can be recommended in clinical practice? *Current Opinion in Pulmonary Medicine* 2018; 24(1): 18-23.

[53] Jones, PW. St. George's respiratory questionnaire: MCID. *COPD* 2005; 2(1): 75-79.

[54] Mahler, DA and Witek, TJ. The MCID of the Transition Dyspnea Index is a Total Score of One Unit. *COPD: Journal of Chronic Obstructive Pulmonary Disease* 2005; 2(1): 99-103.

[55] Donohue, JF. Minimal Clinically Important Differences in COPD Lung Function. *COPD: Journal of Chronic Obstructive Pulmonary Disease* 2005; 2(1): 111-124.

[56] Zhou, Z, Zhou, A, Zhao, Y *et al*. Evaluating the Clinical COPD Questionnaire: A systematic review. *Respirology* 2017; 22(2): 251-262.

[57] Gupta, N, Pinto, LM, Morogan, A *et al*. The COPD assessment test: a systematic review. *European Respiratory Journal* 2014; 44(4): 873-884.

[58] Bohannon, RW and Glenney, SS. Minimal clinically important difference for change in comfortable gait speed of adults with pathology: A systematic review. *Journal of Evaluation in Clinical Practice* 2014; 20(4): 295-300.

[59] Bohannon, RW and Crouch, PT. Minimal clinically important difference for change in 6-minute walk test distance in adults with pathology: a systematic review. *Journal of Evaluation in Clinical Practice* 2017; 23(2): 377-381.

[60] Binenbaum, Y, Amit, M, Billan, S *et al*. Minimal clinically important differences in quality of life scores of oral cavity and oropharynx cancer patients. *Annals of Surgical Oncology* 2014; 21(8): 2773-2781.

[61] Keurentjes, JC, Van Tol, FR, Fiocco, M *et al*. Minimal clinically important differences in health-related quality of life after total hip or knee replacement: A systematic review. *Bone and Joint Research* 2012; 1(5): 71-77.

[62] Schrover, R, Evans, K, Giugliani, R *et al*. Minimal clinically important difference for the 6-min walk test: literature review and application to Morquio A syndrome. *Orphanet Journal of Rare Diseases* 2017; 12(1): 78.

[63] Doganay Erdogan, B, Leung, YY, Pohl, C *et al*. Minimal Clinically Important Difference as Applied in Rheumatology: An OMERACT Rasch Working Group Systematic Review and Critique. *Journal of Rheumatology* 2016; 43(1): 194-202.

[64] Maltenfort, M and Diaz-Ledezma, C. Statistics In Brief: Minimum Clinically Important Difference – Availability of Reliable Estimates. *Clinical Orthopaedics and Related Research* 2017; 475(4): 933-946.

[65] Moher, D, Liberati, A, Tetzlaff, J *et al*. The PRISMA Group Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLoS Medicine* 2009; 6(7): e1000097.

[66] Alma, H, de Jong, C, Tsiligianni, IG *et al*. Measuring clinically relevant change in health status for patients with COPD: a systematic review and meta-analysis. PROSPERO 2015 CRD42015023221.
Available from https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42015023221

[67] Higgins, JPT and Green, S (editors). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. *The Cochrane Collaboration*, 2011. Available from http://handbook.cochrane.org.

[68] Terwee, CB, Mokkink, CB, Knol, DL *et al*. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. *Quality of Life Research* 2012; 21(4): 651-657.

[69] Kamper, SJ, Maher, C and Mackay, G. Global Rating of Change Scales: A Review of Strengths and Weaknesses and Considerations for Design. *Journal of Manual & Manipulative Therapy* 2009; 17(3): 163-170.

[70] Alma, H, de Jong, C, Jelusic, D *et al*. T. Health status instruments for patients with COPD in pulmonary rehabilitation: defining a minimal clinically important difference. *npj Primary Care Respiratory Medicine* 2016; 26: 16041.

[71] Tsai, CL, Hodder, RV, Page, JH *et al.* The short-form chronic respiratory disease questionnaire was a valid, reliable, and responsive quality-of-life instrument in acute exacerbations of chronic obstructive pulmonary disease. *Journal of Clinical Epidemiology* 2008; 61(5): 489-497.

[72] Dodd, JW, Hogg, L, Nolan, J *et al.* The COPD assessment test (CAT): response to pulmonary rehabilitation. A multicenter, prospective study. *Thorax* 2011; 66(5): 425-429.

[73] Jones PW, Harding, G, Wiklung, I *et al.* Tests of the Responsiveness of the COPD Assessment Test Following Acute Exacerbation and Pulmonary Rehabilitation. *Chest* 2012; 142(1): 134-140.

[74] Kocks, JW, Tuinenga, MG, Uil, SM *et al.* Health status measurement in COPD: the minimal clinically important difference of the clinical COPD questionnaire. *Respiratory Research* 2006; 7: 62.

[75] Kon, SS, Canavan, JL, Jones, SE *et al.* Minimum clinically important difference for the COPD Assessment Test: a prospective analysis. *Lancet Respiratory Medicine* 2014; 2(3): 195-203.

[76] Kon, SS, Dilaver, D, Mittal, M *et al.* The Clinical COPD Questionnaire: response to pulmonary rehabilitation and minimal clinically important difference. *Thorax* 2014; 69(9): 793-798.

[77] Kulich, K, Keininger, DL, Tiplady, B, Banerji, D. Symptoms and impact of COPD assessed by an electronic diary in patients with moderate-to-severe COPD: psychometric results from the SHINE study. *International Journal of Chronic Obstructive Pulmonary Disease* 2015; 10: 79-94.

[78] Nolan, CM, Longworth, L, Lord, J *et al.* The EQ-5D-5L health status questionnaire in COPD: validity, responsiveness and minimum important difference. *Thorax* 2016; 71(6): 493-500.

[79] Perez, T, Arnould, B, Grosbois, JM *et al.* Validity, reliability, and responsiveness of a new short Visual Simplified Respiratory Questionnaire (VSRQ) for health-related quality of life assessment in chronic obstructive pulmonary disease. *International Journal of Chronic Obstructive Pulmonary Disease* 2009; 4: 9-18.

[80] Redelmeier, DA, Guyatt, GH and Goldstein, RS. Assessing the Minimal Important Difference in Symptoms: A Comparison of Two Techniques. *Journal of Clinical Epidemiology* 1996; 49(11): 1215-1219.

[81] Schünemann, HJ, Griffith, L, Jaeschke, R *et al.* Evaluation of the minimal important difference for the feeling thermometer and the St. George's Respiratory Questionnaire in patients with chronic airflow obstruction. *Journal of Clinical Epidemiology* 2003; 56(12): 1170-1176.

[82] Smid, DE, Franssen, FM, Houben-Wilke, S *et al.* Responsiveness and MCID Estimates for CAT, CCQ, and HADS in Patients With COPD Undergoing Pulmonary Rehabilitation: A Prospective Analysis. *Journal of the American Medical Directors Association* 2017; 18(1): 53-58.

[83] Tsiligianni, IG, van der Molen, T, Moraitaki, D *et al.* Assessing health status in COPD. A head-to-head comparison between the COPD assessment test (CAT) and the clinical COPD questionnaire (CCQ). *BMC Pulmonary Medicine* 2012, 12: 20.

[84] Van Stel, HF, Maillé, AR, Colland, VT *et al.* Interpretation of change and longitudinal validity of the quality of life for respiratory illness questionnaire (QoLRIQ) in inpatient pulmonary rehabilitation. *Quality of Life Research* 2003; 12(2): 133-145.

[85] Walters, SJ and Brazier, JE. What is the relationship between the minimally important difference and health state utility values? The case of the SF-6D. *Health and Quality of Life Outcomes* 2003; 1: 4.

[86] Walters, SJ and Brazier, JE. Comparison of the minimally important difference for two health state utility measures: EQ-5D and SF-6D. *Quality of Life Research* 2005; 14(6): 1523-1532.

[87] Welling, JB, Hartman, JE, Ten Hacken, NH *et al.* The minimal important difference for the St George's Respiratory Questionnaire in patients with severe COPD. *European Respiratory Journal* 2015; 46(6): 1598-1604.

[88] Wyrwich, KW, Metz, SM, Kroenke, K *et al.* Measuring patient and clinician perspectives to evaluate change in health related quality of life among patients with chronic obstructive pulmonary disease. *Journal of General Internal Medicine* 2007; 22(2): 161-170.

[89] Zanini, A, Aiello, M, Adamo, D *et al.* Estimation of minimal clinically important difference in EQ-5D visual analog scale score after pulmonary rehabilitation in subjects with COPD. *Respiratory Care* 2015; 60(1): 88-95.

[90] Jones, PW, Harding, G, Berry, P *et al.* Development and first validation of the COPD Assessment Test. *European Respiratory Journal* 2009; 34(3): 648-654.

[91] Van der Molen, T, Willemse, BWM, Schokker, S *et al.* Development, validity and responsiveness of the Clinical COPD Questionnaire. *Health and Quality of Life Outcomes* 2003, 1: 13.

[92] Guyatt, GH, Berman, LB, Townsend, M *et al.* A measure of quality of life for clinical trials in chronic lung disease. *Thorax* 1987; 42(10): 773-778.

[93] EuroQol G. EuroQol – a new facility for the measurement of health-related quality of life. *Health Policy* 1990; 16(3): 199-208.

[94] Herdman, M, Gudex, C, Lloyd, A *et al*. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Quality of Life Research* 2011; 20(10): 1727-1736.

[95] Ware, J, Kosinski, M and Dewey, J. How to Score Version Two of the SF-36 Health Survey. Lincoln, RI: QualityMetric Inc.: 2000.

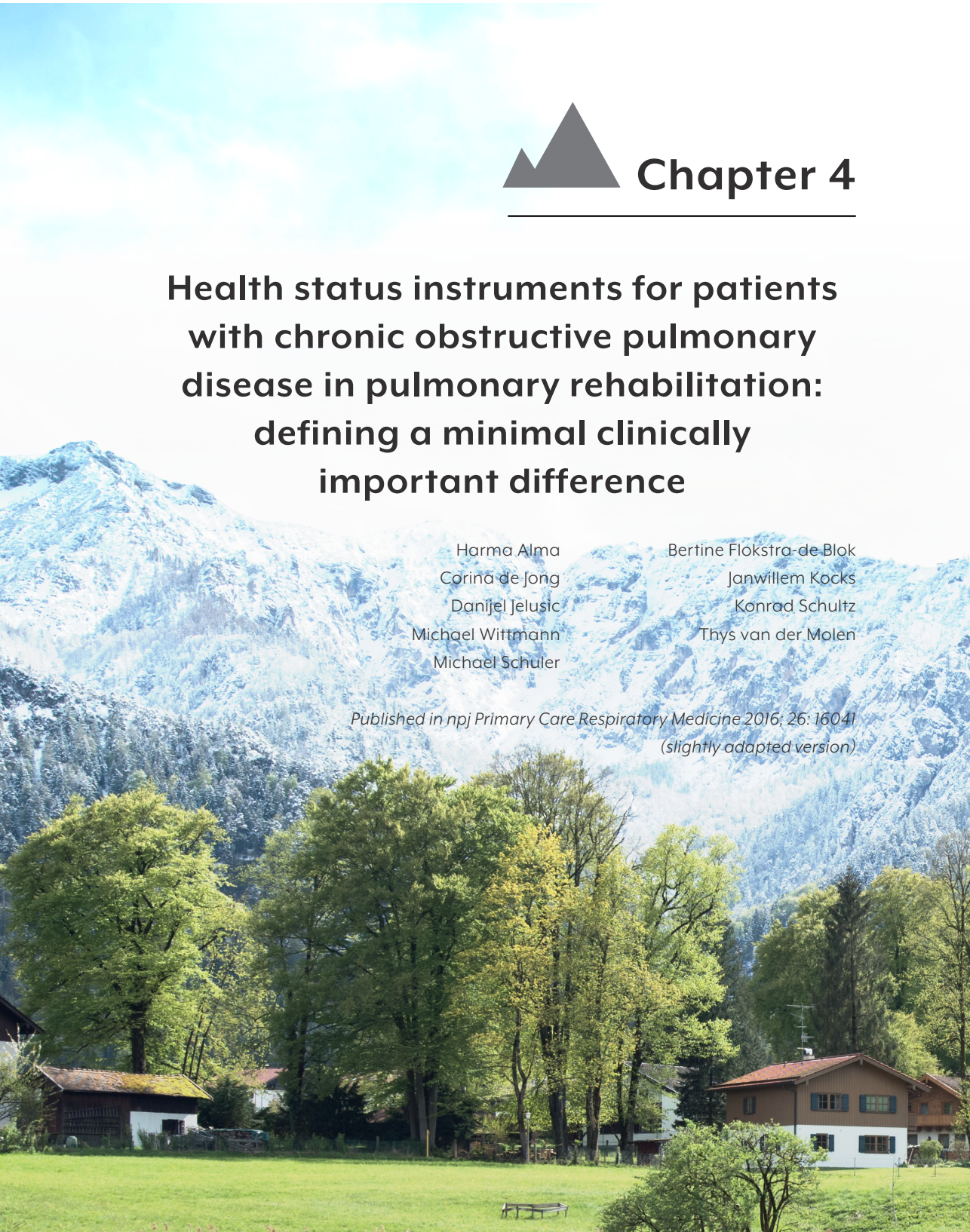
[96] Jones, PW, Quirck, FH and Baveystock, CM. The St George's Respiratory Questionnaire. *Respiratory Medicine* 1991; 85 (Suppl B): 25-31.

[97] Cave, J, Atkinson, L, Tsiligianni, IG *et al*. Assessment of COPD wellness tools for use in primary care: an IPCRG initiative. *International Journal of COPD* 2012; 7: 447-456.

[98] Alma, H, de Jong, C, Jelusic, D *et al*. Assessing health status over time: impact of recall period and anchor question on the minimal clinically important difference of copd health status tools. *Health and Quality of Life Outcomes* 2018; 16(1): 130.

[99] Yoo, JY, Kim, YS, Kim, SS *et al*. Factors affecting the trajectory of health-related quality of life in COPD patients. *International Journal of Tuberculosis and Lung Disease* 2015; 20(6): 738-746.

[100] Sundh, J, Montgomery, S, Hasselgren, M *et al*. Change in health status in COPD: a seven-year follow-up cohort study. *npj Primary Care Respiratory Medicine* 2016; 26: 16073.



Chapter 4

Health status instruments for patients with chronic obstructive pulmonary disease in pulmonary rehabilitation: defining a minimal clinically important difference

Harma Alma
Corina de Jong
Danijel Jelusic
Michael Wittmann
Michael Schuler
Bertine Flokstra-de Blok
Janwillem Kocks
Konrad Schultz
Thys van der Molen

Published in npj Primary Care Respiratory Medicine 2016; 26: 16041
(slightly adapted version)

4.1 Abstract

4.1.1 Background

The minimal clinically important difference (MCID) defines to what extent change on a health status instrument is clinically relevant, which aids scientists and physicians in interpreting therapy effects. This is the first study that aimed to establish the MCID of the COPD Assessment Test (CAT), the Clinical COPD Questionnaire (CCQ) and the St. George's Respiratory Questionnaire (SGRQ) in the same pulmonary rehabilitation (PR) population using multiple approaches.

4.1.2 Methods

In total, 451 patients with chronic obstructive pulmonary disease (COPD) participated in a 3-week PR programme (58 years, 65% male, 43 pack years, global initiative for obstructive lung disease (GOLD) grades II/III/IV 50/39/11%). Techniques used to assess the MCID were anchor-based approaches, including patient-referencing, criterion-referencing and questionnaire-referencing; and the distribution-based methods standard error of measurement (SEM), 1.96SEM and half standard deviation (0.5SD).

4.1.3 Results

Patient- and criterion-referencing led to MCID estimates of -3.12 and -2.96 (CAT); -0.62 and -0.56 (CCQ); and -9.28 and -8.40 (SGRQ). Negative changes represented improvement. Questionnaire-referencing suggested MCID ranges of -3.08 to -1.46 (CAT), -0.61 to -0.28 (CCQ), and -9.47 to -6.86 (SGRQ). The SEM, 1.96SEM and 0.5SD were 3.28, 6.43 and 2.80 (CAT); 0.29, 0.56 and 0.46 (CCQ); 5.20, 10.19 and 6.06 (SGRQ). Pooled estimates were -3.29 (CAT), -0.52 (CCQ) and -7.91 (SGRQ) for improvement. MCID estimates differed by the method used.

4.1.4 Discussion and conclusions

Pooled estimates suggest clinically relevant improvements needing to be at least -3.00 on the CAT, -0.40 on the CCQ and -7.00 on the SGRQ for patients with moderate to very severe COPD. The MCIDs of the CAT and SGRQ in the literature might be too low, leading to overestimation of the interpretation of treatment effects for patients with COPD.

4.2 Background

Chronic obstructive pulmonary disease (COPD) is the third leading cause of death [1]. Spirometry is required to make the diagnosis [2]. Its parameter forced expiratory volume in one second (FEV_1) has however a weak correlation with symptoms and disease impact, which are factors captured by health status instruments [3]. Health status has become an important goal in the management of COPD [2]. Multiple instruments exist that measure health status with the COPD Assessment Test (CAT), Clinical COPD Questionnaire (CCQ) and St. George's Respiratory Questionnaire (SGRQ) most frequently used [4-6]. These tools are important in assessing treatment effectiveness. Therefore, clinically relevant change as outcome of the questionnaires has become pivotal. The minimal clinically important difference (MCID) is a parameter that assesses clinically relevant change. It is defined as "*the smallest difference in score, which patients perceive as beneficial and which would mandate a change in the patient's management*"[7].

Multiple methods for determining the MCID exist, clustered into anchor- and distribution-based approaches [8-10]. Anchor-based approaches require change in health status to be compared with another measure of clinical change, such as a global rating of change (GRC) assessment (*patient-referencing*); the appearance of health events in the time of change (*criterion-referencing*); and/or a related instrument with a known MCID (*questionnaire-referencing*) [8]. Distribution-based methods require comparison of change with a statistical measure of variability of this change such as the standard error of measurement (SEM) or the half standard deviation (0.5SD) [8, 11-12]. Anchor-based methods are preferred as they convey clinical significance, yet distribution-based approaches are quicker to use [9-10]. A golden standard has not been defined. Different methods will lead to a range of estimates [8, 10, 13]. A pledge has been made for an overall body of evidence to agree upon an MCID, or to use multiple MCIDs in practice [8, 10,13-14]. This body of evidence should consist of relevant patient-reported anchors and clinical trial data [10, 15]. However, selecting appropriate anchors is problematic, since this commonly used method is highly dependent on the correlation between instruments (preferably ≥0.50), as well as the accuracy of the anchor instrument's MCID [10, 15].

Existing evidence for the MCID of the CCQ suggests a value of 0.40, which is equivalent to 7% of the scale (range: 0-6) [5, 16-19]. The anchor-based methods patient-referencing, criterion-referencing and questionnaire-referencing with the SGRQ, CAT and Chronic Respiratory Questionnaire (CRQ) as anchors, were separately applied in a Dutch prednisolone trial following acute exacerbation, in pulmonary rehabilitation (PR) for both COPD and non-COPD patients, and in a Greek primary and secondary care population [16-18]. The SEM and 0.5SD techniques were applied too [17-19]. None of the

studies combined all of the approaches in the same population of patients with COPD. The MCID for the domain scores on the CCQ has not been established either. The MCID of the CAT was summarised as 2 points, equivalent to 5% of the scale (range: 0-40) [4, 19-21]. Both anchor- and distribution-based techniques were applied in a PR setting, for acute exacerbation patients with COPD and for outpatients. Criterion-referencing has not been specifically applied for the CAT, nor have all methods been applied simultaneously.

The MCID for the SGRQ is set at 4 points, which is 4% of the scale (range: 0-100) [6, 22-23]. Expert-based ratings, patient-referencing, criterion-referencing, and the use of the 6 minute walking distance (6MWD) and the CRQ as anchors have been applied in various studies on patients with asthma or COPD. These studies are from many years ago; therefore a recent study on patients with severe COPD, who underwent bronchoscopic lung volume reduction (BLVR), claimed the MCID of the SGRQ to be >7 points [24]. It used FEV_1, 6MWD and residual volume (RV) as anchors combined with distribution-based methods. Estimates on the MCID of the SGRQ seem inconsistent. None of the methods have been applied at once, nor has the MCID of the domain scores of the SGRQ been investigated.

The MCIDs of health status tools are necessary for physicians and researchers to evaluate therapy results and clinical trials. Expanding the body of evidence for the MCID remains of major importance. This study is the first to investigate the MCID of the CAT, CCQ and SGRQ simultaneously in PR using the largest array of methods. It examines the impact of using anchor- and distribution-based methods to determine an instrument's MCID. The domain scores on the CCQ and SGRQ are investigated for their MCIDs as well, which is a new development.

4.3 Methods

4.3.1 Study subjects

This study is a secondary analysis of a subsample from the routine inspiratory muscle training within COPD rehabilitation (RIMTCORE) real-life randomised controlled trial *(#DRKS00004609)* in the Klinik Bad Reichenhall, Center for Rehabilitation, Pulmonology and Orthopedics in Germany [25]. Patients with COPD global intiative for obstructive lung disease (GOLD) grades II-IV, ≥18 years, who gave informed consent, were included between February 2013 and July 2014. Exclusion criteria were lack of linguistic or cognitive abilities to fill out questionnaires; hypercapnic respiratory failure with a PaCO2 >50 mmHg at rest or indication for intermittent noninvasive ventilation; contra-indications for inspiratory muscle training (IMT) (e.g. a history of recent lung surgery, fresh pulmonary embolism, history of

recurrent spontaneous pneumothorax); and other severe comorbid diseases that conferred significantly greater morbidity than COPD (e.g. active cancer without successfully completed curative therapy). Patients participated in an intensive 3-week full-day inpatient rehabilitation programme. The therapy components were tailored to the patients' individual needs and included endurance and strength training, patient education, respiratory physiotherapy, psychological support, tobacco cessation and dietary counselling. The RIMTCORE trial was approved by the Ethik-Kommission der Bayerischen Landesärztekammer (*#12107*) and registered in the German clinical trial register.

4.3.2 Study design and data collection

For the current MCID study, completed data were analysed at pre- (T0) and post-inpatient rehabilitation (T1) from a subset of participants without other respiratory comorbidities (e.g. bronchiectasis, asthma, history of bronchial carcinoma, sarcoidosis and tuberculosis) or alpha-1-antitrypsin deficiency. On all measurement occasions, parameters collected were CAT (no recall period), CCQ (weekly version) and SGRQ (monthly version). The CAT is an 8-item unidimensional scale with item scores ranging from 0 to 5 (0: no impairment, 5: maximum impairment) and a total score of 0 to 40 [4]. The CCQ consists of 10 items divided over 3 subdomains. Item scores range from 0 to 6 (0: no impairment, 6: maximum impairment), with the total score derived from adding up item scores and dividing this by 10 [5]. The SGRQ has 50 items divided over 3 domains with a total score of 0 to 100 (0: no impairment, 100: maximum impairment) [6]. A GRC anchor question ranging from -7 to +7 was issued at T1, which required patients to assess their global health in relation to COPD compared with T0. Patient characteristics, post-bronchodilator spirometry, 6MWD and exacerbations in the 12 months before PR were available too.

4.3.3 Determining the MCID: Anchor-based approaches

Patient-referencing: Changes on the CAT, CCQ and SGRQ were categorised according to the GRC score. Scores of 0 and ±1 represented no or hardly any change; scores of ±2 and ±3 were considered a minimal clinically relevant change; scores of ±4 and ±5 were considered a moderate change; and scores of ±6 and ±7 were considered a major change, as exemplified by Juniper *et al.* [26]. The MCID was established by calculating the mean health status change score of the patients with a minimal clinically relevant change on the GRC (i.e. scores of ±2 and ±3).

Criterion-referencing: The health event exacerbation during PR was used as an anchor, which was defined as *"worsening of COPD symptoms requiring at least treatment with oral corticosteroids and antibiotics"*. The difference in baseline score between patients experiencing an exacerbation and those without represented the MCID.

Questionnaire-referencing: Change in one instrument was anchored against change in the other two instruments, as performed prior [17, 20]. Correlations between change scores were assessed, needing to be ≥0.30 (preferably ≥0.50) to be eligible as anchor [10]. The MCID of the anchor from the literature was used as reference (CAT=2.00, CCQ=0.40 and SGRQ=4.00) [16-17, 20, 22]. First, scatter plots and regression analysis with the anchor change score as the independent variable, were produced. Second, the mean was calculated for patients achieving or failing the anchor's MCID. Last, receiver operating characteristics (ROC) curves were plotted to identify the best change in health status to discriminate between those achieving the anchor's MCID and those failing to achieve it [8]. This process resulted in three estimates per anchor. The steps were repeated if the MCID estimates derived from patient-referencing, criterion-referencing and the distribution-based methods were different compared with the literature.

4.3.4 Determining the MCID: Distribution-based approaches

The SEM seeks correlation between single standard error units and established MCID estimates [12]. It is calculated as: SEM = $\sigma_x \sqrt{1-r_{xx}}$, with: r_{xx} = the intraclass correlation coefficient (ICC) and σ_x = standard deviation baseline. Both the SEM and 1.96SEM were calculated, since there is no consensus on which represents the MCID best. The 0.5SD was determined as an equivalent of the MCID [8, 11].

4.3.5 Data analysis

Data analysis was performed using SPSS 20.0 (IBM, Chicago, USA). Descriptive data were evaluated at T0 for frequencies and percentage, or mean and standard deviation. Health status data at T0 and T1 were evaluated with mean and standard deviations, and tested for significance of change with paired t-tests or Wilcoxon signed-rank tests depending on normality of distribution. Negative change represented improvement on the health status instruments. Health status scores of CAT, CCQ and SGRQ were checked for floor- and ceiling effects defined as more than 15% of participating patients scoring in the bottom and top 10% of the maximum scale range.

Patient-referencing approach: Correlations between GRC and CAT, CCQ or SGRQ were assessed using Pearson's or Spearman's coefficients depending on normality. Participants were classified according to GRC score. Significance of change was calculated with paired t-tests or Wilcoxon signed-rank tests depending on normality.

Criterion-referencing approach: The difference in baseline score between patients with and without an exacerbation during PR was evaluated using independent t-tests or Mann-Whitney U tests depending on normality.

Questionnaire-referencing: Correlations were assessed using Pearson's or Spearman's coefficients depending on normality. First, scatter plots and regression analysis were performed with the anchor variable as the independent variable. Next, mean change scores of the CAT, CCQ and SGRQ were calculated for those achieving or failing the suggested MCID of the anchors. ROC curves were plotted with the anchor's MCID as the dichotomizing variable. The optimal value was selected with specificity and sensitivity preferably both ≥0.70, favouring sensitivity.

Distribution-based methods: The SEM, 1.96SEM and 0.5SD of the change for each instrument were calculated. ICC values were obtained from the literature: 0.80 (CAT), 0.94 (CCQ) and 0.91 (SGRQ) [4-6].

Pooled MCID estimates: The mean estimates for the CAT, CCQ and SGRQ derived from patient-referencing, criterion-referencing, questionnaire referencing, the SEM and 0.5SD were multiplied with a factor 1/5 each to calculate a pooled average. Domain scores were averaged based on patient- and criterion-referencing results.

4.4 Results

4.4.1 Patient characteristics and health status

In total, 611 patients participated in the RIMTCORE trial among whom 50 discontinued the study [25]. Out of the remaining participants, 451 met the inclusion criteria for the current MCID analysis. Mean age was 57.87±6.56 years, 65% male and 50/39/11% GOLD grades II/III/IV (*Table 1*).

CAT, CCQ and SGRQ were normally distributed at T0 and T1; change scores were normally distributed for CAT, CCQ and SGRQ symptom scores. Floor- and ceiling effects were negligible except for the CCQ mental domain. There were no missing health status questionnaires at T0. There were four missing participants for the SGRQ, one for the CCQ and two for the CAT at T1. Pair-wise deletion was applied. Mean baseline scores were 20.23±7.33 (CAT), 2.86±1.17 (CCQ) and 50.69±17.33 (SGRQ) with significant improvements after PR of respectively -3.11 (95% confidence interval (CI) -3.63 to -2.59), -0.58 (95%CI -0.67 to -0.50), and -9.04 (95%CI -10.17 to -7.92) (*Table 2*).

Table 1: Baseline characteristics

Variable	Baseline
Age (years)[a]	57.87 ± 6.56
BMI[a]	26.82 ± 6.56
Gender (male)[b]	293 (65.0)
FEV_1%pred[a]	50.40 ± 15.11
GOLD II[b]	227 (50.3)
GOLD III[b]	176 (39.0)
GOLD IV[b]	48 (10.6)
Smoking pack years[a]	42.61 ± 23.47
Never smokers[b]	6 (1.3)
Active smokers[b]	179 (39.7)
Ex-smokers[b]	266 (59.0)
Retired[b]	74 (16.4)
If not retired, unable to work[b]	159 (35.3)
Patients with ≥1 exacerbation during 12 months prior to PR[b]	353 (78.4)

[a] Data expressed as mean ± SD.
[b] Data expressed as frequencies (% of total patients).
N = 451

Abbreviations: BMI, body mass index; FEV_1%pred, forced expiratory volume in one second % predicted; GOLD, global initiative for chronic obstructive lung disease; PR, pulmonary rehabilitation; SD, standard deviation.

4.4.2 MCID: patient-referencing

The GRC score was missing for one patient. Correlations between GRC and health status instruments were significant with r = 0.23 (CAT), 0.29 (CCQ), and 0.30 (SGRQ). In total, 12 patients showed deterioration on the GRC (GRC ≤-2). No or hardly any improvement (GRC= 0, or +1) was experienced by 21.7% (n=98). Minimal improvement (GRC= +2 and +3) was seen in 43.5% of patients (n=196), whereas moderate (GRC= +4 and +5) and major improvement (GRC= +6 and +7) represented respectively 27.3% (n=123) and 5.5% (n=25) (*Table 3*). At the threshold for minimal clinically relevant improvement (GRC= +2 or +3) mean CAT, CCQ and SGRQ change scores were respectively -3.12 (95%CI -3.86 to -2.37), -0.56 (95%CI -0.68 to -0.44) and -8.40 (95% CI -10.07 to -6.73). Mean improvements for these patients on the CCQ domains were -0.55 (symptoms), -0.55 (functional), and -0.58 (mental); and for the SGRQ domains -13.12 (symptoms), -5.98 (activity), and -8.24 (impact).

Table 2: Health status outcomes of pulmonary rehabilitation

Instrument	Baseline[a]	Change[a]	95% CI[b]
CAT			
Total	20.23 ± 7.33	-3.11 ± 5.59	-3.63 to -2.59
CCQ			
Symptoms	2.87 ± 1.24	-0.59 ± 1.16	-0.70 to -0.48
Functional	2.86 ± 1.34	-0.56 ± 1.00	-0.65 to -0.46
Mental	2.86 ± 1.74	-0.62 ± 1.49	-0.76 to -0.48
Total	2.86 ± 1.17	-0.58 ± 0.92	-0.67 to -0.50
SGRQ			
Symptoms	63.66 ± 21.77	-14.22 ± 21.69	-16.24 to -12.21
Activities	63.58 ± 19.82	-6.71 ± 13.44	-7.96 to -5.47
Impact	39.21 ± 18.81	-8.78 ± 13.95	-10.08 to -7.49
Total	50.69 ± 17.33	-9.04 ± 12.11	-10.17 to -7.92
6MWD (meters)	427.73 ± 110.18	80.19 ± 54.72	75.01 to 85.37

Negative change represents improvement on the CAT, CCQ and SGRQ.
[a] Data expressed as mean ± SD.
[b] Paired t-tests applied for normally distributed data, Wilcoxon signed rank tests for non-parametric data.
All tests significant at level $P < 0.05$ comparing pre- and post-pulmonary rehabilitation scores.
N=451

Abbreviations: 6MWD, six minute walking distance; 95%CI, 95% confidence interval; CAT, COPD Assessment Test; CCQ, Clinical COPD Questionnaire; COPD, chronic obstructive pulmonary disease; SD, standard deviation; SGRQ, St. George's Respiratory Questionnaire.

4.4.3 MCID: criterion-referencing

During PR, 10% of patients (n=45) experienced an exacerbation. There were no missing data. Mean differences between both groups at baseline were -2.96 (95%CI -5.20 to -0.71) for CAT, -0.62 (95%CI -0.98 to -0.27) for CCQ and -9.28 (95%CI -14.56 to -3.99) for SGRQ, with significantly higher scores for patients with an exacerbation. Significant domain differences were -0.47, -0.67 and -0.86 for the respective CCQ symptoms, functional and mental domains; and -10.61 and -9.93 for the SGRQ activity and impact domains.

4.4.4 MCID: questionnaire-referencing

Significant correlations between total change scores were 0.63 (SGRQ versus CCQ), 0.54 (SGRQ versus CAT) and 0.59 (CCQ versus CAT) (*Supplementary material 4.7.1*). Using the original anchor estimates from the literature (CAT=-2.00, CCQ=-0.40 and SGRQ=-4.00) the various questionnaire-referencing results including 95%CI resulted in the following ranges: -3.00 to -2.14 (CCQ as anchor) and -3.00 to -1.46 (SGRQ as anchor) for CAT; -0.53 to -0.42 (CAT as anchor) and -0.50 to -0.28 (SGRQ as anchor) for CCQ; -8.30 to -6.86 (CCQ as anchor) and -8.48 to -6.98 (CAT as anchor) for SGRQ (*Table 4, Online Supplementary Figures 1-6 at npj Primary Care Respiratory Medicine*).

Table 3: MCID patient-referencing results

Instrument	No/hardly any improvement (GRC -1, 0 or +1) N= 98		Minimal improvement (GRC +2 or +3) N= 196		Moderate improvement (GRC +4 or +5) N= 123		Major improvement (GRC +6 or +7) N= 25	
	Δ[a]	95%CI[b]	Δ[a]	95%CI[b]	Δ[a]	95%CI[b]	Δ[a]	95%CI[b]
CAT								
Total	-2.05	-3.13 to -0.98	-3.12	-3.86 to -2.37	-3.67	-4.70 to -2.67	-6.44	-8.99 to -3.89
CCQ								
Symptoms	-0.32	-0.54 to -0.10	-0.55	-0.70 to -0.40	-0.76	-0.96 to -0.56	-1.48	-1.97 to -0.99
Functional	-0.27	-0.47 to -0.07	-0.55	-0.68 to -0.43	-0.78	-0.98 to -0.58	-0.97	-1.31 to -0.63
Mental	-0.53	-0.84 to -0.22	-0.58	-0.78 to -0.38	-0.67	-0.94 to -0.39	-1.34	-1.84 to -0.84
Total	-0.34	-0.52 to -0.15	-0.56	-0.68 to -0.44	-0.75	-0.92 to -0.58	-1.25	-1.54 to -0.96
SGRQ								
Symptoms	-7.03	-10.86 to -3.19	-13.12	-16.05 to -10.19	-19.91	-23.92 to -15.90	-30.62	-38.40 to -22.84
Activities	-3.03	-5.28 to -0.78	-5.98	-7.88 to -4.08	-10.33	-12.78 to -7.87	-12.66	-18.69 to -6.62
Impact	-6.72	-9.57 to -3.86	-8.24	-10.14 to -6.33	-10.32	-12.77 to -7.87	-17.90	-23.64 to -12.16
Total	-5.57	-7.79 to -3.35	-8.40	-10.07 to -6.73	-11.83	-14.00 to -9.66	-18.50	-22.81 to -14.18

Negative change represents improvement for all instruments.

[a] Data reported as mean change scores.

[b] Paired t-tests applied to normally distributed variables and Wilcoxon signed rank tests were used for not normally distributed data. Data reported as 95%CI.

All change scores significant at level P < 0.05.

Abbreviations: Δ, change score; 95%CI, 95% confidence interval; CAT, COPD Assessment Test; CCQ, Clinical COPD Questionnaire; COPD, chronic obstructive pulmonary disease; GRC, global rating of change; MCID, minimal clinically important difference; SGRQ, St. George's Respiratory Questionnaire.

Using averaged estimates from the other MCID approaches in this study (CAT=-3.00, CCQ=-0.50 and SGRQ=-7.00) the results including 95%CI were: -3.08 to -2.54 (CCQ as anchor) and -3.00 to -2.32 (SGRQ as anchor) for CAT; -0.61 to -0.53 (CAT as anchor) and -0.60 to -0.44 (SGRQ as anchor) for CCQ; -8.90 to -7.79 (CCQ as anchor) and -9.47 to -8.00 (CAT as anchor) for SGRQ (*Table 4, Online Supplementary Figures 1-6 at npj Primary Care Respiratory Medicine*).

Online Supplementary Figures 1-6

4.4.5 MCID: Distribution-based approach

The SEM for the CAT, CCQ and SGRQ was 3.28, 0.29 and 5.20, respectively; the 1.96SEM was 6.43, 0.56 and 10.19, respectively; and the 0.5SD was 2.80, 0.46 and 6.06, respectively.

Table 4: MCID questionnaire-referencing results

	Anchor CCQ=-0.40	Anchor CCQ=-0.50	Anchor CAT=-2	Anchor CAT=-3	Anchor SGRQ=-4	Anchor SGRQ=-7
Regression Analysis						
CAT	-2.45 (-2.77 to -2.14)	-2.81 (-3.08 to -2.54)	-	-	-1.86 (-2.27 to -1.46)	-2.61 (-2.91 to -2.32)
CCQ	-	-	-0.48 (-0.53 to -0.42	-0.57 (-0.61 to -0.53)	-0.34 (-0.40 to -0.28)	-0.48 (-0.53 to -0.44)
SGRQ	-7.51 (-8.16 to -6.86)	-8.35 (-8.90 to -7.79)	-7.73 (-8.48 to -6.98)	-8.89 (-9.47 to -8.31)	-	-
Failing/Achieving						
CAT	-2.74	-2.82	-	-	-2.45	-2.86
CCQ	-	-	-0.48	-0.56	-0.46	-0.53
SGRQ	-8.14	-8.36	-7.78	-8.69	-	-
ROC Curves						
CAT	-3.00	-3.00	-	-	-3.00	-3.00
	AUC 0.768 Sens 0.726 Spec 0.656	AUC 0.770 Sens 0.729 Spec 0.668			AUC 0.722 Sens 0.727 Spec 0.605	AUC 0.737 Sens 0.705 Spec 0.650
CCQ	-	-	-0.50	-0.60	-0.50	-0.60
			AUC 0.767 Sens 0.701 Spec 0.706	AUC 0.771 Sens 0.716 Spec 0.710	AUC 0.796 Sens 0.750 Spec 0.714	AUC 0.802 Sens 0.763 Spec 0.730
SGRQ	-8.30	-8.63	-7.50	-8.00	-	-
	AUC 0.817 Sens 0.777 Spec 0.703	AUC 0.816 Sens 0.787 Spec 0.702	AUC 0.719 Sens 0.659 Spec 0.656	AUC 0.745 Sens 0.673 Spec 0.681		

Data expressed as estimates (95%CI).
N=451

Abbreviations: 95%CI, 95% confidence interval; AUC, area under the curve; CAT, COPD Assessment Test; CCQ, Clinical COPD Questionnaire; COPD, chronic obstructive pulmonary disease; ROC, receiver operating characteristics; Sens, sensitivity; Spec, specificity; SGRQ, St. George's Respiratory Questionnaire.

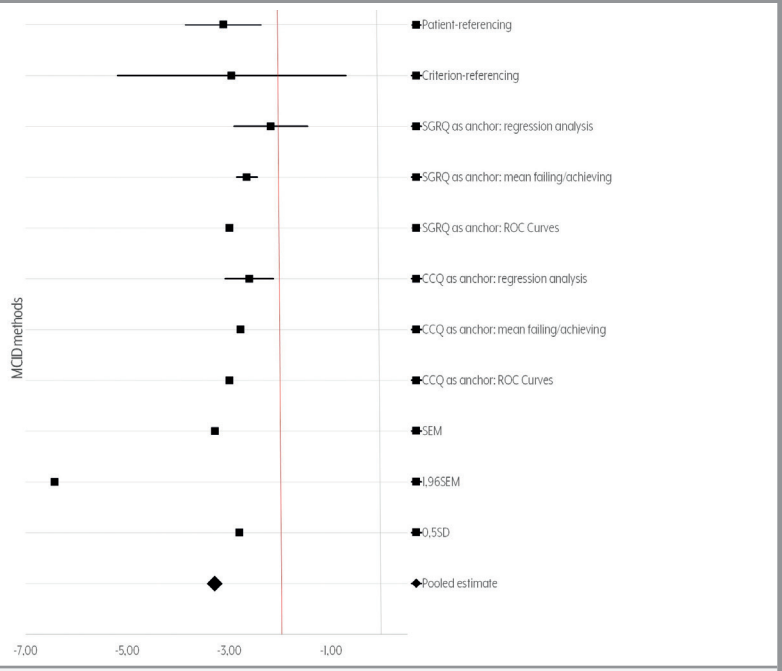
4.4.6 Pooled MCID estimates

The weighted MCID estimates were -3.29 (CAT), -0.52 (CCQ), and -7.91 (SGRQ). Results for the domains were -0.51 (symptoms), -0.61 (functional status) and -0.72 (mental) for CCQ; and -13.12 (symptoms), -8.30 (activities) and -9.09 (impact) for SGRQ. Results from all approaches are visualised in *Figures 1-3 (Supplementary material 4.7.2).*

4.4.7 Power analysis

Post-hoc analysis demonstrated the power of the study to be over 90% based upon the number of cases (N=451), alpha 0.05 and the effect sizes for CAT (0.43), CCQ (0.50) and SGRQ (0.53).

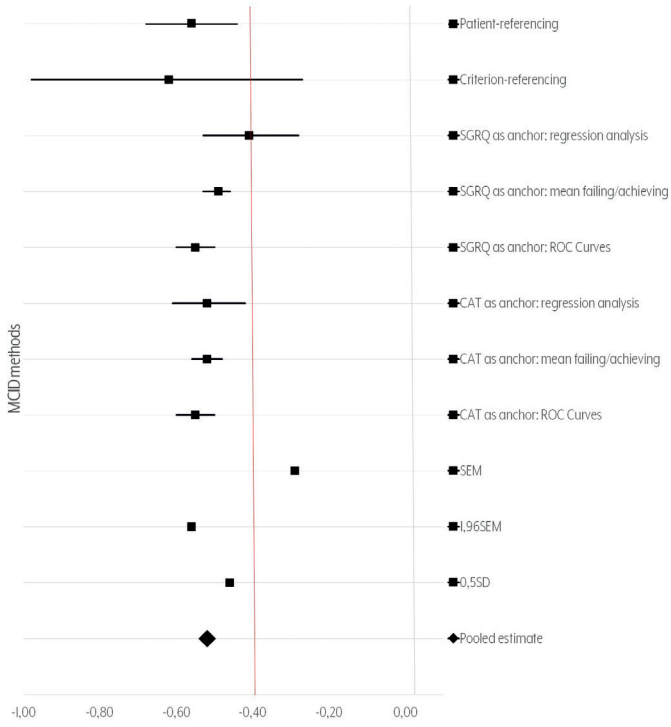
Figure 1: Summary plot of the MCID estimates of the CAT



The horizontal plots represent the MCID estimates derived in this study, classified per method. Where appropriate the estimates include the 95% confidence interval. The red vertical line resembles the MCID estimate obtained from the literature.

Abbreviations: CAT, COPD Assessment Test; CCQ, Clinical COPD Questionnaire; COPD, chronic obstructive pulmonary disease; MCID, minimal clinically important difference; ROC, receiver operating characteristics; SD, standard deviation; SEM, standard error of measurement; SGRQ, St. George's Respiratory Questionnaire.

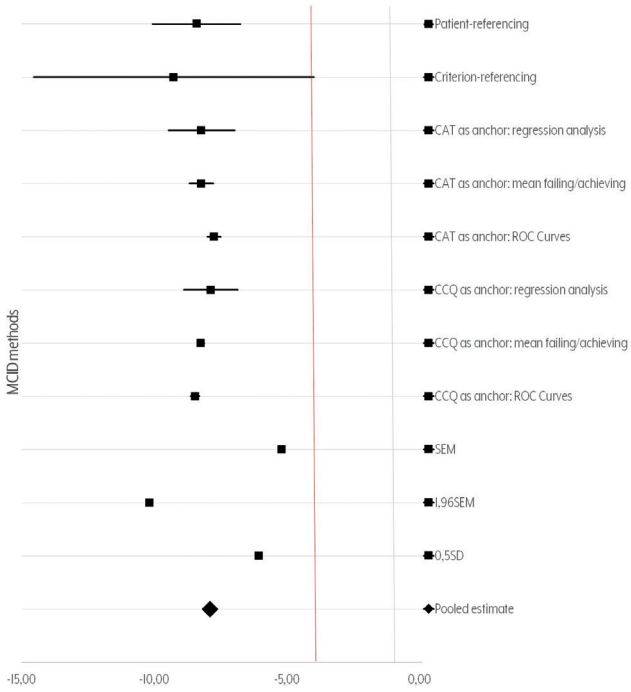
Figure 2: Summary plot of the MCID estimates of the CCQ



The horizontal plots represent the MCID estimates derived in this study, classified per method. Where appropriate the estimates include the 95% confidence interval. The red vertical line resembles the MCID estimate obtained from the literature.

Abbreviations: CAT, COPD Assessment Test; CCQ, Clinical COPD Questionnaire; COPD, chronic obstructive pulmonary disease; MCID, minimal clinically important difference; ROC, receiver operating characteristics; SD, standard deviation; SEM, standard error of measurement; SGRQ, St. George's Respiratory Questionnaire.

Figure 3: Summary plot of the MCID estimates of the SGRQ



The horizontal plots represent the MCID estimates derived in this study, classified per method. Where appropriate the estimates include the 95% confidence interval. The red vertical line resembles the MCID estimate obtained from the literature.

Abbreviations: CAT, COPD Assessment Test; CCQ, Clinical COPD Questionnaire; COPD, chronic obstructive pulmonary disease; MCID, minimal clinically important difference; ROC, receiver operating characteristics; SD, standard deviation; SEM, standard error of measurement; SGRQ, St. George's Respiratory Questionnaire.

4.5 Discussion

4.5.1 Main findings

The use of anchor- and distribution-based methods in this study resulted in MCID ranges of -6.43 to -1.46 for CAT, -0.62 to -0.28 for CCQ, and -10.19 to -5.20 for SGRQ. The pooled MCID estimates derived from the various methods in this study (-3.29 for CAT, -0.52 for CCQ, and -7.91 for SGRQ) are similar or slightly higher compared with the literature. In general, results from the patient-referencing method were somewhat comparable to criterion- and questionnaire-referencing. The adjusted MCID cutpoints of the anchors (SGRQ=-7.00, CCQ=-0.50 and CAT=-3.00) in the questionnaire-referencing approach had slightly better correspondence with patient- and criterion-referencing. The distribution-based method 0.5SD was best comparable to anchor-based results. The SEM was inconsistent, and the 1.96SEM was much more conservative for CAT and SGRQ, although not for CCQ.

4.5.2 Interpretation of findings

Almost all MCID estimates including the pooled estimate (-3.29) for the CAT were higher than the suggested 2 points in the literature. Patient-referencing resulted in a much higher estimate compared with Dodd *et al.* and Kon *et al.*, who used both a GRC with just 5 choice options [20-21]. Preferably, more answering categories on the anchor question should be used to provide the full-spectrum of answers [27]. Criterion-referencing in our study was comparable to the other anchor-based approaches. This method has not been performed for the CAT before. Questionnaire-referencing resulted in ranges of -3.00 to -1.46 (SGRQ as anchor) and -3.08 to -2.14 (CCQ as anchor). The use of the original MCID of the SGRQ as anchor provided the lowest estimates for the CAT, just as for CCQ. Possibly the MCID of the SGRQ is not as solid as claimed. The ranges found in our study matched results from Kon *et al.* [20], but they are higher than the 2 points summarised. Earlier, CAT has been mapped to the SGRQ, resulting in an MCID of 1.60 [5, 21]. It seems that this derived from multiplication of the SGRQ MCID with a factor 40/100, which is rather unusual. A similar exercise for CCQ would result in an MCID of 0.24, far below current estimates. The distribution-based methods SEM and 0.5SD matched results from anchor-based approaches. The 1.96SEM is much higher and lacked correspondence.

The pooled MCID estimate for the CCQ in our study (-0.52) is slightly higher than the literature estimate. Patient- and criterion-referencing in our study used comparable methodology [16, 18], but resulted in more conservative estimates. Differences in participants' age, baseline CCQ score and period of measurement possibly influenced this. Our study included younger patients with more severe baseline scores. Furthermore, exacerbations might not be a minor event for the included patients. Questionnaire-referencing resulted

in ranges of -0.60 to -0.28 (SGRQ as anchor) and -0.61 to-0.42 (CAT as anchor), which to some extent matched results of Kon *et al.* for COPD [17], as well as of Canavan *et al.* for other respiratory diseases [18]. With regard to the distribution-based methods, the 1.96SEM and 0.5SD were best comparable to results from anchor-based approaches, but slightly higher than the results of Kocks *et al.* [16]. The 0.5SD matched earlier results [17-18]. Domain MCID scores were approximately equivalent to MCID estimates of the total CCQ score, although the mental domain was higher, possibly because of floor- and ceiling effects in the current study.

All MCID estimates for SGRQ were larger than the 4 points from the literature [22]. MCIDs of the domain scores on the SGRQ (except for symptoms) seemed comparable to the estimate for the total score. The suggested literature MCID of the SGRQ originates from patient-referencing in two studies, featuring the use of salmeterol in COPD and nedocromil sodium in asthma [28-29], in which a limited 5-point GRC scale was used to review therapy effects [23]. Osman *et al.* report the results of criterion-referencing comparing SGRQ scores between patients re-admitted within 12 months and those, who were not, resulting in an MCID estimate of 4.80 [30]. The results in the current study are nearly double the original MCID estimates, which date back to the nineties. Differences in study setting, age of patients, time period of measurement, and different health event criterion may have influenced this large difference. Poor methodologic quality of the patient-referencing approach might be another explanation.

Questionnaire-referencing provided ranges of -9.47 to -6.98 (CAT as anchor) and -8.90 to -6.86 (CCQ as anchor), which was somewhat comparable to patient- and criterion-referencing results. MCID estimates from the adjusted questionnaire-referencing approach were slightly higher and better comparable with the other anchor-based approaches. Upon development of the SGRQ, a hypothesised multivariate model estimated a 6% mean difference on the 6MWD (22m) to be equivalent to 4 points on the SGRQ [6]. However, nowadays the MCID of the 6MWD is considered to be doubled [31]. Schünemann *et al.* found a change of (-)3.05 on the SGRQ to match the MCID of the CRQ dyspnoea domain [32]; however this is only a measure for dyspnoea and not the complete health status concept. It could have underestimated the MCID of the SGRQ severely. Recently, a study by Welling *et al.* suggested the MCID of the SGRQ to be over 7 points for patients with severe COPD undergoing BLVR using both anchor-based and distribution-based methods [24]. Our current study includes moderate to very severe patients with COPD in PR. The results overlap one another using different anchors.

The SEM and the 0.5SD for the SGRQ are both lower than the anchor-based results, whereas the 1.96SEM was higher and lacked correspondence too. Tsiligianni *et al.* calculated the 1.96SEM of the SGRQ to be 4.84, which is substantially lower than all estimates here [19]. Jones argued that distribution-based methods were not applicable to the SGRQ, since they lack agreement with anchor-based approaches, and determined a Standard Error of the Estimate (1.3) and 0.5SD (8.4) based upon averaging data from 11 studies [22]. This pooled 0.5SD matched with results in our study.

4.5.3 Strengths and limitations of this study

This study applied for the first time multiple approaches to determine the MCID of the CAT, CCQ and SGRQ simultaneously in one strong dataset of patients with COPD in PR. It is also the first study to include estimates of the possible MCIDs for domain scores of the CCQ and SGRQ as well. Estimates are valid for improvement and for patients with moderate to very severe COPD (GOLD grades II-IV). No patients with mild COPD (GOLD grade I) were included in this study. During PR too few patients deteriorated (n=12) to analyse the MCID for deterioration. MCIDs for improvement and deterioration may differ though [13].

Our current study has applied an ambiguous anchor-based method of questionnaire-referencing. However, this approach is widely used and accepted elsewhere to estimate another instrument's MCID [17-18, 20]. The pooled thresholds for clinically relevant change of the CCQ, CAT and SGRQ in our study (CCQ=-0.52, CAT=-3.29 and SGRQ=-7.91) seem different from values reported in the literature (CCQ=-0.40, CAT=-2.00 and SGRQ=-4.00). This has had impact on the questionnaire-referencing results. Averaged MCID estimates from patient-referencing, criterion-referencing and distribution-based methods were included therefore in the analysis as cutoff values for the anchor's MCID. The revised MCIDs of the anchor questionnaires had better correspondence with results from these other approaches. It highlights that careful selection of anchors should be considered.

A limitation of the current study is that the correlations between both CAT, CCQ SGRQ, and the GRC were below the preferred lower limit to be appropriate as anchor ($r \geq 0.30$ and preferably ≥ 0.50). Other studies using a GRC seldom published correlation coefficients, making it unclear whether this problem is widespread. Another limitation of this study is that the PR period was 3 weeks, whereas the SGRQ has a recall period of 1 month and the CCQ has a recall period of 1 week. A study by Meguro *et al.* compared the shorter SGRQ-C without recall period with the original SGRQ with a specified 4-week recall period [33]. No differences in scores between both questionnaires were observed. We therefore expect little influence of the recall period on our MCID results. Last, exacerbations during PR were used as criterion to estimate the MCID. The estimates from this approach are higher for

CCQ and SGRQ compared with other methods. This might indicate that exacerbations were not a minor clinically relevant event for patients.

4.5.4 Implications for future research, policy and practice

Our study demonstrated that the existing MCIDs of the SGRQ and possibly of the CAT are set too low in current practice. Using a low threshold could lead to overestimation of treatment effects. Patients currently considered to experience clinically relevant change as a result of treatment may in fact not experience this. On the other hand, a more conservative cutoff point may not approve therapy, although benefits for the patient do exist. Even though the current study adds to the body of evidence, the analysis is based on one patient group only, where many would be preferred. More studies are necessary to build a more complete body of evidence and understanding, preferably with the full scope of approaches and in different patient groups. These should also further investigate whether the MCID for the domain scores for the CCQ and SGRQ is comparable to the total score estimate. The quest in finding the gold standard for the MCID for health status tools for COPD must continue.

4.5.5 Conclusions

The current study suggests that improvements need to be at least 3 points on the CAT, 0.40 points on the CCQ, and 7 points on the SGRQ to be considered clinically relevant for patients with moderate to very severe COPD. The MCID for domain scores on the SGRQ and CCQ could be equivalent to these thresholds.

4.6 Declarations

4.6.1 Funding

The main RIMTCORE trial (*#DRKS00004609*) concerning the effects of pulmonary rehabilitation was funded by the Deutsche Rentenversicherung. The current study regarding the MCID of the CAT, CCQ and SGRQ received financial support from the Junior Scientific Masterclass as part of the University of Groningen.

4.6.2 Acknowledgements

We are grateful to the Junior Scientific Masterclass of the University of Groningen, who financially supported the research position of the first author.

4.6.3 Authors' contributions and consent

Konrad Schultz, Michael Wittmann, Danijel Jelusic and Michael Schuler planned the RIMTCORE study design regarding the effects of inspiratory muscle training during pulmonary rehabilitation, and were responsible for data collection. Harma Alma, Corina de Jong and Thys van der Molen designed the current study regarding the MCID of the CAT, CCQ and SGRQ. Harma Alma and Corina de Jong performed the statistical analysis. Harma Alma wrote the first draft, while Corina de Jong, Bertine Flokstra-de Blok, Janwillem Kocks and Thys van der Molen actively participated in the review process. Thys van der Molen supervised and participated in different steps of the study, as well as in writing. All authors participated in various steps in the study, edited the manuscript and gave their approval for submission.

4.6.4 Competing interests

Konrad Schultz received lecture fees from Boehringer Ingelheim, AstraZeneca, Berlin Chemie, Novartis, Chiesi, Mundipharma, Takeda, GSK and MSD, all outside the submitted work. Janwillem Kocks reports personal fees from Novartis; personal fees from Boehringer Ingelheim; grants and personal fees from GSK; grants from Stichting Zorgdraad; personal fees from IPCRG; personal fees from Springer Media; and travel arrangements from Chiesi BV, GlaxoSmithKline BV, and IPCRG, all outside the submitted work. Janwillem Kocks is an Associate Editor of npj Primary Care Respiratory Medicine, but was not involved in the editorial review of, nor the decision to publish, this article. Thys van der Molen developed the CCQ and holds the copyright. The remaining authors declare no conflict of interest.

4.7 Supplementary material

4.7.1 Health status correlations

Table 1: Relevant correlations between health status questionnaires

	CAT	SGRQ			
CCQ		Symptoms	Activity	Impact	Total
Symptoms	0.47	0.51*	0.33	0.38	0.50*
Functional	0.45	0.39	0.43	0.47	0.55*
Mental	0.36	0.27	0.29	0.43	0.43
Total	0.59*	0.50*	0.44	0.54*	0.63*
CAT	-	0.40	0.33	0.41	0.54*

Data expressed as correlation coefficients.
All correlations significant at level P < 0.05.
* Correlations were ≥0.50.
N= 451

Abbreviations: CAT, COPD Assessment Test; CCQ, Clinical COPD Questionnaire; COPD, chronic obstructive pulmonary disease; SGRQ, St. George's Respiratory Questionnaire.

4.7.2 Summary MCID results

Table 2: Summary of all MCID Results

MCID approaches	CAT	CCQ	SGRQ
Estimate from literature	(-)2.00	(-)0.40	(-)4.00
Patient-referencing	-3.12	-0.56	-8.40
		-0.55 symptoms	-13.12 symptoms
		-0.55 functional	-5.98 activity
		-0.58 mental	-8.24 impact
Criterion-referencing	-2.96	-0.62	-9.28
		-0.47 symptoms	- symptoms
		-0.67 functional	-10.61 activity
		-0.86 mental	-9.93 impact
Questionnaire-referencing			
SGRQ=4 and SGRQ=7 as anchor			
Scatter plots / regression	-2.91 to -1.46	-0.53 to -0.28	
Mean failing/achieving estimate	-2.86 to -2.45	-0.53 to -0.46	
ROC Curves	-3.00 to -3.00	-0.60 to -0.50	
CAT=2 and CAT=3 as anchor			
Scatter plots / regression		-0.61 to -0.42	-9.47 to -6.98
Mean failing/achieving estimate		-0.56 to -0.48	-8.69 to -7.78
ROC Curves		-0.60 to -0.50	-8.00 to -7.50
CCQ=0.40 and CCQ=0.50 as anchor			
Scatter plots / regression	-3.08 to -2.14		-8.90 to -6.86
Mean failing/achieving estimate	-2.82 to -2.74		-8.36 to -8.14
ROC Curves	-3.00 to -3.00		-8.63 to -8.30
Distribution-based approaches			
SEM	3.28	0.29	5.20
1.96SEM	6.43	0.56	10.19
0.5SD	2.80	0.46	6.06
Pooled estimate	-3.29	-0.52	-7.91
		-0.51 symptoms	-13.12 symptoms
		-0.61 functional	-8.30 activity
		-0.72 mental	-9.09 impact

N=451

Abbreviations: 0.5SD, half standard deviation; CAT, COPD Assessment Test; CCQ, Clinical COPD Questionnaire; COPD, chronic obstructive pulmonary disease; MCID, minimal clinically important difference; ROC, receiver operating characteristics; SD, standard deviation; SEM, standard error of measurement; SGRQ, St. George's Respiratory Questionnaire.

4.8 References

[1] Lozano, R, Naghavi, M, Foreman, K *et al*. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet* 2012; 380(9859): 2095-2128.

[2] Global initiative for Chronic Obstructive Lung Disease (GOLD). Global Strategy for the Diagnosis, Management and Prevention of COPD. 2016. Available at: http://www.goldcopd.org/uploads/users/files/GOLD_Report%20 2016.pdf. Accessed 1st of July 2016.

[3] Westwood, M, Bourbeau, J, Jones, PW *et al*. Relationship between FEV1 change and patient-reported outcomes in randomized trials of inhaled bronchodilators for stable COPD: a systematic review. *Respiratory Research* 2011; 12: 40.

[4] Jones, PW, Harding, G, Berry, P *et al*. Development and first validation of the COPD Assessment Test. *European Respiratory Journal* 2009; 34(3): 648-654.

[5] Van der Molen, T, Willemse, BWM, Schokker, S *et al*. Development, validity and responsiveness of the Clinical COPD Questionnaire. *Health and Quality of Life Outcomes* 2003, 1: 13.

[6] Jones, PW, Quirck, FH and Baveystock, CM. The St George's Respiratory Questionnaire. *Respiratory Medicine* 1991; 85 (Suppl B): 25-31.

[7] Jaeschke, R, Singer, J and Guyatt, GH. Measurement of health status. Ascertaining the Minimal Clinically Important Difference. *Controlled Clinical Trials* 1989; 10(4): 407-415.

[8] Copay, AG, Subach, BR, Glassman, SD *et al*. Understanding the minimum clinically important difference: a review of concepts and methods. *The Spine Journal* 2007; 7(5): 541-546.

[9] Lydick, E and Epstein, RS. Interpretation of quality of life changes. *Quality of Life Research* 1993; 2(3): 221-226.

[10] Revicki, D, Hays, RD, Cella, D *et al*. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *Journal of Clinical Epidemiology* 2008; 61(2): 102-109.

[11] Norman, GR, Sloan, JA and Wyrwich, KW. Interpretation of Changes in Health-related Quality of Life. The Remarkable Universality of Half a Standard Deviation. *Medical Care* 2003; 41(5): 582-592.

[12] Wyrwich, KW, Nienaber, NA, Tierney, WM *et al*. Linking Clinical Relevance and Statistical Significance in Evaluating Intra-Individual Changes in Health-Related Quality of Life. *Medical Care* 1999; 37(5): 469-478.

[13] Beaton, DE, Boers, M and Wells, GA. Many faces of the minimal clinically important difference (MCID): a literature review and directions for future research. *Current Opinion in Rheumatology* 2002; 14(2): 109-114.

[14] Curtis, JR and Patrick, DL. The assessment of health status among patients with COPD. *European Respiratory Journal Supplement* 2003; 41: 36s-45s.

[15] Revicki, DA, Cella, D, Hays, RD *et al*. Responsiveness and minimal important differences for patient reported outcomes. *Health and Quality of Life Outcomes* 2006; 4: 70.

[16] Kocks, JW, Tuinenga, MG, Uil, SM *et al*. Health status measurement in COPD: the minimal clinically important difference of the clinical COPD questionnaire. *Respiratory Research* 2006; 7: 62.

[17] Kon, SS, Dilaver, D, Mittal, M *et al*. The Clinical COPD Questionnaire: response to pulmonary rehabilitation and minimal clinically important difference. *Thorax* 2014; 69(9): 793-798.

[18] Canavan, JL, Dilaver, D, Clark, AL *et al*. Clinical COPD Questionnaire in patients with chronic respiratory disease. *Respirology* 2014; 19(7): 1006-1012.

[19] Tsiligianni, IG, van der Molen, T, Moraitaki, D *et al*. Assessing health status in COPD. A head-to-head comparison between the COPD assessment test (CAT) and the clinical COPD questionnaire (CCQ). *BMC Pulmonary Medicine* 2012, 12: 20.

[20] Kon, SS, Canavan, JL, Jones, SE *et al*. Minimum clinically important difference for the COPD Assessment Test: a prospective analysis. *Lancet Respiratory Medicine* 2014; 2(3): 195-203.

[21] Dodd, JW, Hogg, L, Nolan, J *et al*. The COPD assessment test (CAT): response to pulmonary rehabilitation. A multicenter, prospective study. *Thorax* 2011; 66(5): 425-429.

[22] Jones, PW. St. George's respiratory questionnaire: MCID. *COPD* 2005; 2(1): 75-79.

[23] Jones, PW. Interpreting thresholds for a clinically significant change in health status in asthma and COPD. *European Respiratory Journal* 2002; 19(3): 398-404.

[24] Welling, JB, Hartman, JE, Ten Hacken, NH *et al.* The minimal important difference for the St George's Respiratory Questionnaire in patients with severe COPD. *European Respiratory Journal* 2015; 46(6): 1598-1604.

[25] Schultz, K, Jelusic, D, Wittmann, M *et al.* Inspiratory muscle training does not improve clinical outcomes in 3-week COPD rehabilitation: results from a randomised controlled trial. *European Respiratory Journal* 2018; 51(1): 1702000.

[26] Juniper, EF, Guyatt, GH, Willan, A *et al.* Determining a Minimal Important Change in a Disease-Specific Quality of Life Questionnaire. *Journal of Clinical Epidemiology* 1994; 47(1): 81-87.

[27] Kamper, SJ, Maher, C and Mackay, G. Global Rating of Change Scales: A Review of Strengths and Weaknesses and Considerations for Design. *Journal of Manual & Manipulative Therapy* 2009; 17(3): 163-170.

[28] Jones, PW. Quality of life, symptoms and pulmonary function in asthma: long-term treatment with nedocromil sodium examined in a controlled multicentre trial. Nedocromil Sodium Quality of Life Study Group. *European Respiratory Journal* 1994; 7(1): 55-62.

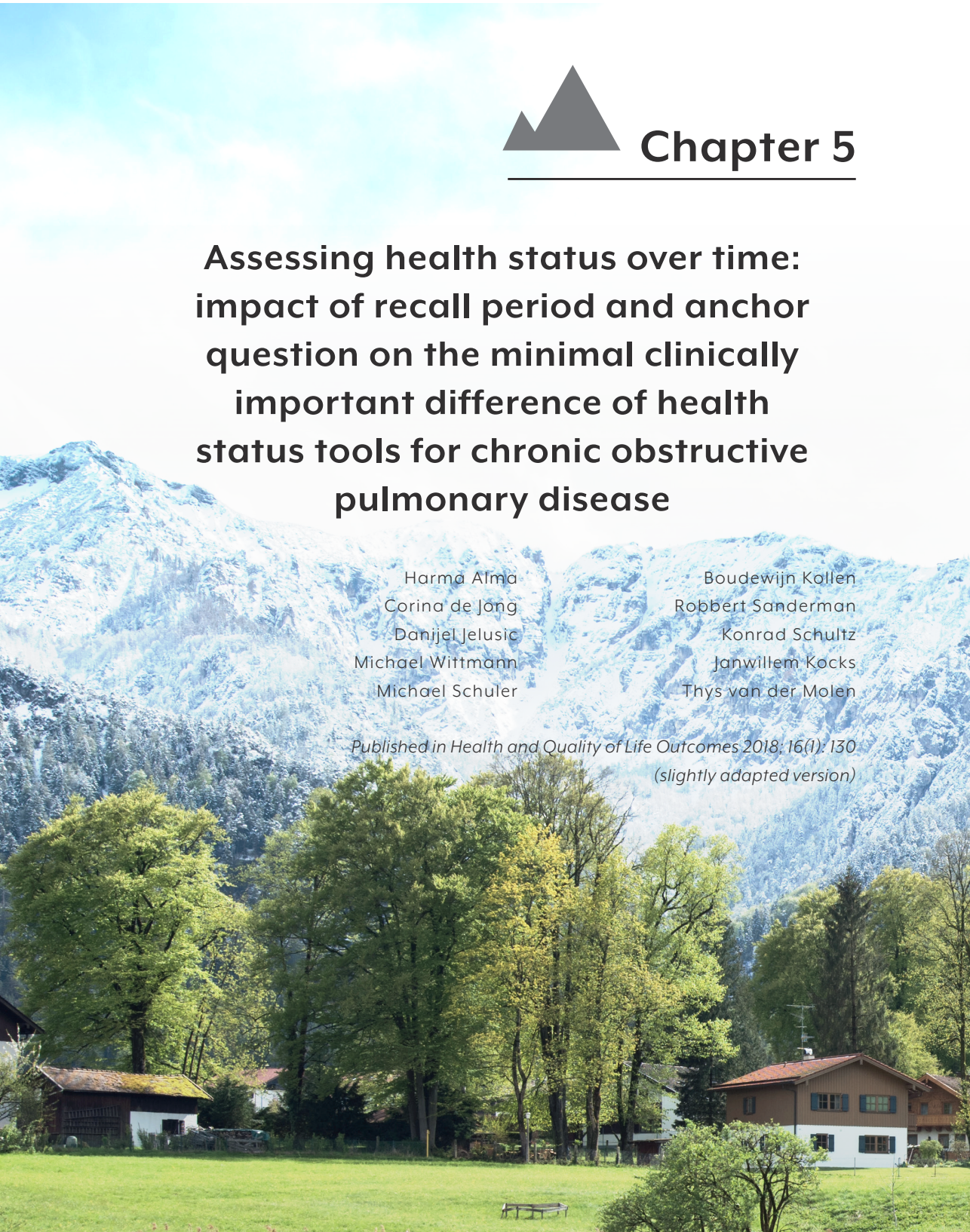
[29] Jones, PW and Bosh, TK. Quality of life changes in COPD patients treated with salmeterol. *American Journal of Respiratory Critical Care Medicine* 1997; 155(4): 1283-1289.

[30] Osman, IM, Godden, DJ, Friend, JA *et al.* Quality of life and hospital re-admission in patients with chronic obstructive pulmonary disease. *Thorax* 1997; 52(1): 67-71.

[31] Redelmeier, DA, Bayoumi, AM, Goldstein, RS, Guyatt, GH. Interpreting small differences in functional status: the Six Minute Walk test in chronic lung disease patients. *American Journal of Respiratory Critical Care Medicine* 1997; 155(4): 1278-1282.

[32] Schünemann, HJ, Griffith, L, Jaeschke, R *et al.* Evaluation of the minimal important difference for the feeling thermometer and the St. George's Respiratory Questionnaire in patients with chronic airflow obstruction. *Journal of Clinical Epidemiology* 2003; 56(12): 1170-1176.

[33] Meguro, M, Barley, EA, Spencer, S *et al.* Development and Validation of an Improved, COPD-Specific Version of the St. George Respiratory Questionnaire. *Chest* 2007; 132(2): 456-463.



Chapter 5

Assessing health status over time: impact of recall period and anchor question on the minimal clinically important difference of health status tools for chronic obstructive pulmonary disease

Harma Alma
Corina de Jong
Danijel Jelusic
Michael Wittmann
Michael Schuler
Boudewijn Kollen
Robbert Sanderman
Konrad Schultz
Janwillem Kocks
Thys van der Molen

Published in Health and Quality of Life Outcomes 2018; 16(1): 130
(slightly adapted version)

5.1 Abstract

5.1.1 Background

The minimal clinically important difference (MCID) assesses what change on a measurement tool can be considered minimal clinically relevant. Although the recall period can influence questionnaire scores, it is unclear if it influences the MCID. This study is the first to examine longitudinally the impact of the recall period of an anchor question and its design on the MCID for health status tools for patients with chronic obstructive pulmonary disease (COPD) using the COPD Assessment Test (CAT), Clinical COPD Questionnaire (CCQ) and St. George's Respiratory Questionnaire (SGRQ).

5.1.2 Methods

Patients with moderate to very severe COPD, but without respiratory comorbidities, were recruited during 3-week pulmonary rehabilitation (PR). CAT, CCQ and SGRQ were completed at baseline, discharge, 3, 6, 9 and 12 months. A 15-point global rating of change scale (GRC) was completed at each follow-up. A 5-point GRC was used as a second anchor at 12 months. Mean change scores of a subset of patients indicating a minimal improvement on each of the anchor questions were considered the MCID. The MCID estimates over different time periods were compared with one another by evaluating the degree of overlap of confidence intervals (CI) adjusted for dependency.

5.1.3 Results

In total, 451 patients were included (57.9±6.6 years, 65% male, 50/39/11% global initiative for obstructive lung disease class II/III/IV), of whom 309 completed follow-up. Baseline health status scores were 20.2±7.3 (CAT), 2.9±1.2 (CCQ) and 50.7±17.3 (SGRQ). MCID estimates for improvement ranged -3.1 to -1.4 for CAT, -0.6 to -0.3 for CCQ, and -10.3 to -7.6 for SGRQ. Absolute higher - though not significant - MCIDs were observed for CAT and CCQ directly after PR. Significantly absolute lower MCID estimates were observed for CAT (difference -1.4: CI -2.3 to -0.5) and CCQ (difference -0.2: CI -0.3 to -0.1) using a 5-point GRC.

5.1.4 Discussion and conclusions

The recall period of a 15-point anchor question seemed to have limited impact on the MCID for improvement of the CAT, CCQ and SGRQ during PR; although a 3-week MCID estimate directly after PR intervention might lead to absolute higher values. However, the design of the anchor question was likely to influence the MCID of the CAT and CCQ.

5.2 Background

Health status can be defined as "*the impact of health on a person's ability to perform and derive fulfilment from the activities of daily life*" [1]. Its measurement is a standardised means of quantifying this impact on a patient's daily life, health and wellbeing [1-2]. Multiple general- and disease-specific health status tools have been developed to detect and quantify health status [3-4]. Physiological measures alone do not reflect the full impact of the disease and correlations with health-related quality of life (HRQoL) are often weak [4]. Determining treatment effects requires a parameter that assesses to what extent change on a health status tool can be considered clinically relevant. The minimal clinically important difference (MCID) is used to evaluate this. It has been defined as "*the smallest difference in score, which patients perceive as beneficial and which would mandate a change in the patient's management*" [5]. Observed change should exceed the estimated MCID value in order to be clinically relevant.

MCID estimates can be determined using both anchor- and distribution-based methods [6-8]. A frequently applied anchor-based technique is the use of a reference (*anchor*) question, requiring patients to retrospectively assess their current health state compared to a prior measurement in time or their experienced degree of change over time [6-8]. This anchor question usually consists of multiple ordinal reply options varying from much worse, a little worse, no change, a little better, up to much better [9-10]. The technique may also be referred to as *patient-referencing* [6]. In the literature, several descriptions are used for this kind of anchor question: global rating of change scale (GRC), patient global impression of change, global perceived change, transition rating scale and many more [9-10]. The MCID of a health status instrument can be determined by calculating the mean change score observed for those patients indicating a minimal change (i.e., *little better* or *little worse*) on the anchor question, assuming data being normally distributed [9].

The use of these patient rating scales has pros and cons. Its main strengths are the ease of administration and MCID determination, as well as the involvement of a patient-related clinical anchor [9]. However, it remains unclear over which period of time change on a GRC should be assessed and how many answering options the anchor question should include. When assessing change over a longer period of time, it might be more difficult for the patient to recall their former health state. A longer recall period could result in a different MCID [10]. On the other hand, shorter periods of measurement may not reflect real change. There is no golden standard in defining an instrument's MCID [11].

In chronic obstructive pulmonary disease (COPD) much focus is nowadays on health status measurement [12-13], because spirometry assessment has only a weak to moderate correlation with the patient's wellbeing [14-15]. The COPD Assessment Test (CAT) [16], the Clinical COPD Questionnaire (CCQ) [17], and the St. George's Respiratory Questionnaire (SGRQ) [18] are recommended by the global initiative for chronic obstructive lung disease (GOLD) for the assessment of COPD in order to determine whether a patient is symptomatic and to what extent therapy has been successful [19]. The CAT and CCQ are most applicable in clinical practice, and the SGRQ in scientific research [19-20].

Various studies examined the MCID of the CCQ to be 0.40-0.50 points [21-26], including three studies using an anchor question with recall periods ranging from 2 to 3 days [21], up to 3 weeks [25], and 8 weeks [23]. The MCID of the CAT was estimated to be 2-3 points [24-28], of which three studies used an anchor question with recall periods of 3 weeks [25] and 8 weeks [27-28]. For the SGRQ, the MCID of 4 points is frequently used in clinical trials. However, estimates in the literature range 4-8 points [25, 29-31], of which two studies used patient-referencing techniques with recall periods of 3 weeks [25] and 16 weeks [29, 31]. No studies have investigated the influence of the recall period of the anchor question and the number of its ordinal reply categories upon the MCID of these instruments. Therefore, this study aimed to investigate the impact of the length of the anchor's recall period and the number of reply options on the GRC on the MCID of the most frequently used health status tools CAT, CCQ and SGRQ in patients with COPD recruited from a pulmonary rehabilitation (PR) setting.

5.3 Methods

5.3.1 Study subjects

The routine inspiratory muscle training within COPD rehabilitation (RIMTCORE) study was a real-life randomised controlled trial (German clinical trial *#DRKS00004609*) in the Klinik Bad Reichenhall, Center for Rehabilitation, Pulmonology and Orthopedics in Germany [32]. Patients were included between February 2013 and July 2014. Detailed inclusion- and exclusion criteria have been published elsewhere [25, 32]. This study is a secondary analysis of a subsample including participants with COPD GOLD grades II-IV, aged ≥18 years, who gave informed consent, and who were without respiratory comorbidities (e.g. bronchiectasis, asthma, history of bronchial carcinoma, sarcoidosis, tuberculosis) or alpha-1-antitrypsin deficiency.

5.3.2 Study design and data collection

Patients participated in an intensive 3-week full-day inpatient PR programme tailored to the patient's individual needs including components of physical training, education, smoking cessation, physiotherapy and counselling [25, 32]. Patient characteristics and post-bronchodilator spirometry were collected at baseline and after 3 weeks at the end of PR. Primary parameters collected for this substudy were the CAT (no recall period), CCQ (weekly version) and SGRQ (monthly version) at baseline, discharge and during follow-up measurements at 3, 6, 9 and 12 months. Measurements were taken in the clinic pre- and post-rehabilitation. Patients were blinded to their previous answers during PR. The remaining follow-up questionnaires were sent to the patient's home by regular mail.

The CAT is an 8-item unidimensional scale with item scores ranging from 0 to 5 (0: no impairment; 5: maximum impairment), summing up to a total of maximum 40 points [16]. The CCQ consists of 10 items scoring from 0 to 6 (0: no impairment; 6: maximum impairment) [17]. Domain scores (symptoms, functional status and mental status) and the total questionnaire score can be determined by summing all relevant item scores divided by the number of items. The SGRQ has 50 items divided over the domains symptoms, activities and impact [18]. Scores are calculated using the developers' scoring file. Domain and total SGRQ scores can range from 0 to 100 (0: no impairment, 100: maximum impairment). Scores of the CAT and CCQ were multiplied and standardised into a scale from 0 to 100 to be comparable with SGRQ. All questionnaires were validated and reliable in primary and secondary care, as well as PR for patients with COPD [18, 29, 33-34]. The three instruments are recommended according to the GOLD strategical guidelines [19].

At each follow-up moment a 15-point Likert scale GRC anchor question was scored by the patients requiring assessment of their global health in relation to their COPD compared with the start of PR (*Supplementary material 5.7.1*). Answers were marked on a scale from -7 to +7, ranging from very much worse to very much better and zero representing no change [9]. At 12-months follow-up a 5-point GRC, analogue to the second question of the Short-Form 36 (SF-36), was also scored by the patient (*Supplementary material 5.7.2*) [35]. It required patients to rate their general health compared to one year prior. Patients could assess their status as the same, somewhat better or somewhat worse, or as much better or much worse. Both GRCs are frequently used in MCID research [9]. The term recall period in this sense, refers to the recall period of the GRCs.

5.3.3 Determining the MCID

Scores for CAT, CCQ and SGRQ refer to their total scores. All change scores on the three questionnaires were calculated as the difference between baseline and each respective follow-up measurement. Negative change on these health status tools indicated

improvement and positive change represented deterioration in HRQoL. Changes on these instruments were categorised using the corresponding score on the GRC anchor question. Scores of 0 and ±1 on the 15-point GRC indicated no change; scores of ±2 and ±3 represented a minimal change; scores of ±4 and ±5 were summarised as a moderate change; and scores of ±6 and ±7 indicated a large change [9]. The 5-point GRC resulted in a division of patients as not changed, somewhat better, somewhat worse, much better, or much worse [35]. MCID estimates for the CAT, CCQ and SGRQ total scores were calculated as the mean change scores compared with baseline including the 95% confidence interval (95%CI) of those patients indicating a minimal improvement (+2 and +3) on the GRC at each follow-up measurement, after checking for normality of distribution of the data. In addition to the 15-point Likert GRC scale, the 5-point anchor question was used in a similar way to classify patients as somewhat better. Only patients that indicated an improvement on the GRC were included, since patients tend to get better after intervention and a limited number of patients were expected to deteriorate.

5.3.4 Data analysis

Data analysis was performed using SPSS 23.0 (IBM, Chicago, USA). Descriptive data were evaluated at baseline for either frequencies with percentages (%), mean with standard deviation (SD) or median with range. This was depending on the variable characteristics and/or normality of distribution. CAT, CCQ and SGRQ were evaluated at baseline (T0), at discharge (T1), after 3 months (T2), after 6 months (T3), after 9 months (T4) and after 12 months (T5). Normality of distribution was assessed using histograms combined with skewness and kurtosis results. Values between -1 and +1 were considered indicative for normality. Mean and SDs (or median and range) were calculated for each measurement. Data were checked for floor- and ceiling effects defined as more than 15% of the patients in the lowest and highest 10% of the maximum scale score [36]. All health status change scores were calculated between baseline and each follow-up measurement. These change scores were tested for significance using paired t-tests after verifying normality of distribution. All tests were assessed for significance using the level $P<0.05$.

The MCID determination process included several steps. First, correlations between the GRC anchor questions, and respectively the CAT, CCQ or SGRQ were assessed using Pearson or Spearman correlation coefficients depending on normality of distribution. Correlations (r) needed to be ≥0.30 (preferably ≥0.50) to be eligible as anchor [7]. Next, participants were categorised according to their GRC score at each follow-up measurement. The respective change versus baseline was tested for significance using paired t-tests after checking for normality. Each MCID estimate was calculated as the mean change score compared with baseline including its 95% confidence interval (CI) for those patients indicating a minimal improvement/somewhat better on the GRC for

each follow-up moment. Correspondence between the 15-point and 5-point GRC was analysed using cross tabulations, correlation coefficients and bar charts.

All MCID estimates were tested for significance with one another by determining the degree of overlap of the adjusted CIs. Due to the dependency of the data, the intraclass correlation coefficient (ICC) between follow-up measurement and baseline was calculated and used to construct CIs. Adjusted CIs were calculated based on the ICC between follow-up moment and baseline [37]. The degree of dependency affects the width of the CI required to be able to test for significant differences between the various MCID estimates. Results were visualised in plots. A lack of overlap between the MCID estimates and their respective CIs indicated significant differences between MCIDs. Finally, the MCID estimates and their adjusted CIs from the current study were also compared with the available thresholds from the literature (CAT (-)2.00, CCQ (-)0.40, and SGRQ (-)4.00 points).

5.4 Results

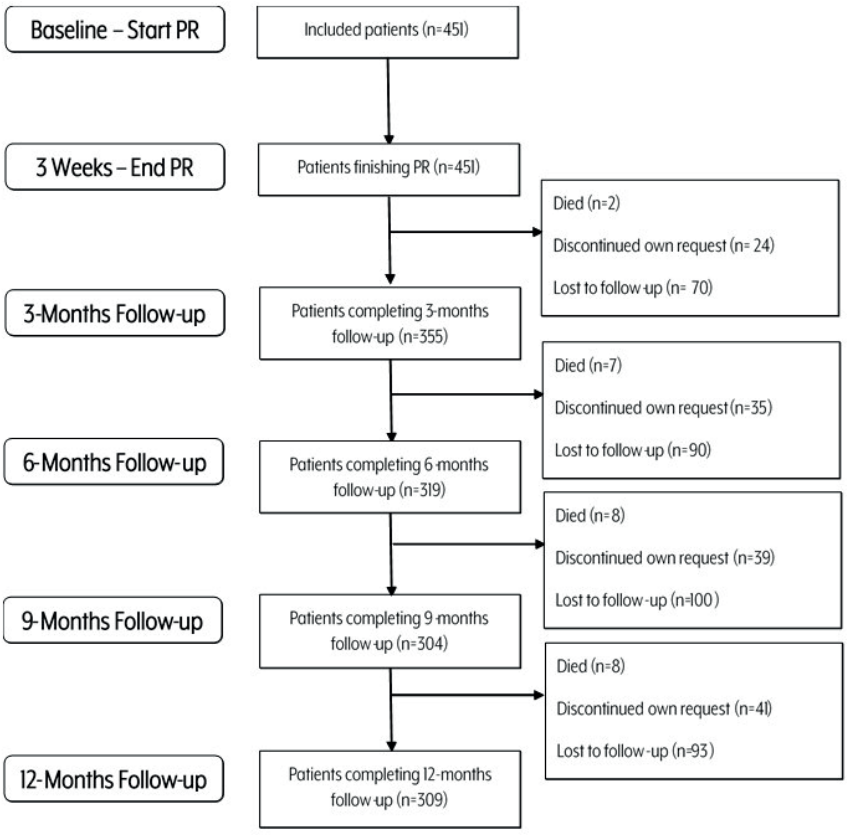
5.4.1 Patient characteristics

This secondary analysis of the RIMTCORE trial included 451 patients [32]. All patients had completed baseline data and at discharge, with the exception for one incomplete CCQ questionnaire, two incomplete CAT questionnaires and four incomplete SGRQ questionnaires at discharge. During follow-up 355 patients had completed data after 3 months; 319 after 6 months; 304 after 9 months; and 309 after 12 months (*Figure 1*). In total, eight patients died during follow-up according to our knowledge, 41 dropped out at own request and a varying number of non-responses at follow-up was present. Mean age was 58 years, 65% was male and had a mean forced expiratory volume in one second % predicted (FEV_1%pred) of 50.4±15.1 (*Table 1*). There were no significant baseline differences between patients completing the 12-months follow-up and those who did not. Full patient characteristics at baseline have been published elsewhere [25].

5.4.2 Health status scores

CAT, CCQ and SGRQ total scores were normally distributed for all measurement moments between T0 and T5. Completed pairs of change scores (follow-up versus baseline) were included only (i.e., *pair-wise deletion*). There were no floor- and ceiling effects observed. There were no significant baseline differences in health status between complete and incomplete follow-up patients (*Table 1*). Mean baseline scores were 20.2±7.3 (CAT), 2.9±1.2 (CCQ) and 50.7±17.3 (SGRQ) (*Table 1*). Mean change after 12 months of follow-up was significant compared with baseline of -0.9 (95%CI -1.7 to -0.1) for CAT; -0.2 (95%CI -0.3 to -0.1) for CCQ; and -3.9 (95%CI -5.7 to -2.2) for SGRQ (*Table 2*).

Figure 1: Consort flow-chart of the number (n) of patients during follow-up



Abbreviations: N, number of patients; PR, pulmonary rehabilitation

Table 1: Baseline patient characteristics

Variable	Baseline all patients n=451	Patients with complete follow-up n=309	Patients with incomplete follow-up at 12 months n=142	Significance
Age (years)[a]	57.9 ± 6.6	58.1 ± 6.5	57.5 ± 6.6	p = 0.39
Gender (male)[b]	293 (65.0)	197 (63.8)	96 (67.6)	p = 0.43
FEV_1%pred[a]	50.4 ± 15.1	50.6 ± 14.9	50.0 ± 15.6	p = 0.72
GOLD II[b]	227 (50.3)	158 (51.1)	69 (48.6)	
GOLD III[b]	176 (39.0)	121 (39.2)	55 (38.7)	p = 0.63
GOLD IV[b]	48 (10.6)	30 (9.7)	18 (12.7)	
Smoking pack years[a]	42.6 ± 23.5	41.1 ± 23.1	45.9 ± 24.0	p = 0.05
CAT[a]	20.2 ± 7.3	20.0 ± 7.6	20.8 ± 6.8	p = 0.28
CCQ Total[a]	2.9 ± 1.2	2.8 ± 1.2	3.0 ± 1.1	p = 0.06
SGRQ Total[a]	50.7 ± 17.3	49.6 ± 17.8	53.0 ± 16.2	p = 0.05

[a] Data expressed as mean ± SD.
[b] Data expressed as frequencies (% of total).
Significance tested with independent t-tests or Chi Square tests at level P < 0.05.

Abbreviations: CAT, COPD Assessment Test; CCQ, Clinical COPD Questionnaire; COPD, chronic obstructive pulmonary disease; FEV_1%pred, forced expiratory volume in one second % predicted; GOLD, global initiative for chronic obstructive lung disease; N, number of patients; SD, standard deviation; SGRQ, St. George's Respiratory Questionnaire.

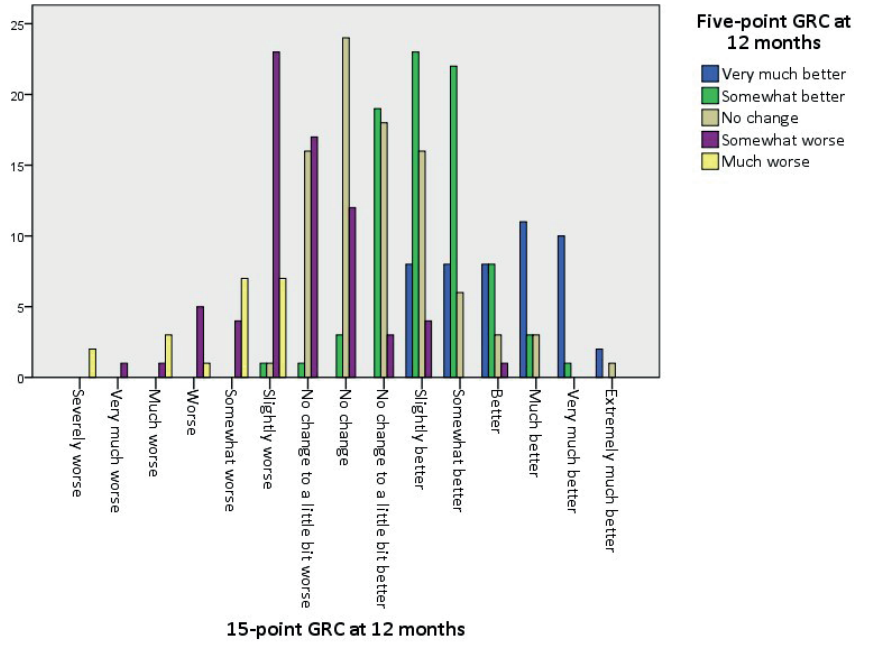
Table 2: Health status change scores

	N	CAT	CAT Standardised	CCQ Total	CCQ Total Standardised	SGRQ Total
Change at discharge (T1)	451	-3.1* (-3.6 to -2.6)	-7.8* (-9.1 to 6.5)	-0.6* (-0.7 to -0.5)	-9.7* (-11.2 to -8.3)	-9.0* (-10.2 to -7.9)
Change after 3 months (T2)	355	-1.4* (-2.2 to -0.7)	-3.6* (-5.4 to -1.8)	-0.3* (-0.4 to -0.2)	-4.3* (-6.2 to -2.5)	-5.4* (-6.9 to -3.8)
Change after 6 months (T3)	319	-0.9* (-1.7 to -0.1)	-2.3* (-4.2 to -0.4)	-0.1 (-0.2 to zero)	-1.8 (-3.8 to +0.2)	-4.9* (-6.5 to -3.2)
Change after 9 months (T4)	304	-1.1* (-1.9 to -0.4)	-2.9* (-4.8 to -0.9)	-0.2* (-0.4 to -0.1)	-3.8* (-5.8 to -1.8)	-5.2* (-6.9 to -3.4)
Change after 12 months (T5)	309	-0.9* (-1.7 to -0.1)	-2.2* (-4.2 to -0.3)	-0.2* (-0.3 to -0.1)	-2.7* (-4.7 to -0.7)	-3.9* (-5.7 to -2.2)

Change scores were calculated in comparison to baseline.
Negative values represent improvement for CAT, CCQ and SGRQ.
Change scores reported as means (95%CI).
CAT and CCQ were standardised into a scale from zero to 100 to be comparable with SGRQ.
* Significance of change scores at level P < 0.05 at T1/T2/T3/T4/T5 compared to baseline T0.

Abbreviations: 95%CI, 95% confidence interval; CAT, COPD Assessment Test; CCQ, Clinical COPD Questionnaire; COPD, chronic obstructive pulmonary disease; N, number of patients; SGRQ, St. George's Respiratory Questionnaire; T0, baseline measurement; T1, measurement at discharge; T2, 3-months follow-up; T3, 6-months follow-up; T4, 9-months follow-up; T5, 12-months follow-up.

Figure 2: Correspondence between the 5- and 15-point global rating of change anchor at 12-months follow-up



Abbreviations: GRC, global rating of change

Table 3: Correlations between health status change scores and the global rating of change anchor questions

	15-point GRC T1	15-point GRC T2	15-point GRC T3	15-point GRC T4	15-point GRC T5	5-point GRC T5
N of Patients	451	355	319	304	309	309
CAT change score	-0.23	-0.33	-0.40	-0.43	-0.41	0.46
CCQ total change score	-0.29	-0.42	-0.44	-0.48	-0.47	0.50*
SGRQ total change score	-0.30	-0.48	-0.51*	-0.58*	-0.54*	0.57*

Data reported as Pearson or Spearman correlation coefficients between health status change scores and the respective anchor GRC.
* Correlations were ≥0.50.

Abbreviations: CAT, COPD Assessment Test; CCQ, Clinical COPD Questionnaire; COPD, chronic obstructive pulmonary disease; GRC, global rating of change; SGRQ, St. George's Respiratory Questionnaire; T1, measurement at discharge; T2, 3-months follow-up; T3, 6-months follow-up; T4, 9-months follow-up; T5, 12-months follow-up.

5.4.3 Minimal clinically important differences for the CAT, CCQ and SGRQ

All change scores and 15-point anchor question scores were normally distributed. The 5-point GRC at 12 months was treated as non-parametric data. At T1, one patient had a missing GRC score. No other GRC scores were missing for T2-T5. Correlations between the 5-/15-point anchor questions and the health status change scores on the CAT, CCQ and SGRQ were all ≥0.30, except for CCQ and CAT at T1 (*Table 3*). The Spearman correlation coefficient between the 5- and 15-point GRC at 12 months was 0.81. The overlap between the 5-point GRC and the 15-point GRC classification at 12-months was 55% based upon a cross-tabulation (*Figure 2*).

A subset of the total patient population, indicated a minimal improvement according to their GRC score. Patients indicating a minimal improvement on the 15-point GRC (scores of +2 or +3) noted significant absolute mean changes between the start of PR and 12 months of follow-up of -2.8 (95%CI -4.2 to -1.4) on the CAT; -0.5 (95%CI -0.7 to -0.3) on the CCQ; and -8.8 (95%CI -11.8 to -5.8) on the SGRQ (*Table 4*). MCID estimates ranged from -3.1 to -2.3 for CAT; -0.6 to -0.4 for CCQ; and from -10.3 to -7.6 for the SGRQ. Mean change scores of those patients feeling somewhat better on the 5-point GRC after 12 months were -1.4 for CAT (95%CI -2.7 to -0.1), -0.3 for CCQ (95%CI -0.5 to -0.2), and -7.7 for SGRQ (95%CI -10.5 to -4.8) (*Table 4*).

Table 4: MCID estimates for minimally improved patients as indicated on the GRC during follow-up

Measurement period	N	CAT	CAT Standardised	CCQ Total	CCQ Total Standardised	SGRQ Total
15-point GRC T1-T0	196	-3.1*± 5.3 (-3.9 to -2.4)	-7.8*± 13.2 (-9.7 to -5.9)	-0.6*± 0.8 (-0.7 to -0.4)	-9.3*± 14.0 (-11.3 to -7.3)	-8.4*± 11.8 (-10.1 to -6.7)
15-point GRC T2-T0	107	-2.7*± 6.4 (-4.0 to -1.5)	-6.9*± 16.1 (-9.9 to -3.8)	-0.4*± 1.0 (-0.6 to -0.2)	-7.3*± 16.8 (-10.5 to -4.0)	-7.6*± 13.8 (-10.2 to -4.9)
15-point GRC T3-T0	96	-2.7*± 6.7 (-4.1 to -1.4)	-6.8*± 16.7 (-10.2 to -3.5)	-0.4*± 1.1 (-0.6 to -0.2)	-7.0*± 18.0 (-10.6 to -3.3)	-9.2*± 14.0 (-12.1 to -6.3)
15-point GRC T4-T0	80	-2.3*± 6.1 (-3.7 to -1.0)	-5.8*± 15.2 (-9.2 to -2.4)	-0.5*± 0.8 (-0.7 to -0.3)	-7.6*± 13.8 (-10.8 to -4.7)	-10.3*± 12.9 (-13.2 to -7.4)
15-point GRC T5-T0	88	-2.8*± 6.7 (-4.2 to -1.4)	-7.0*± 16.7 (-10.5 to -3.5)	-0.5*± 1.0 (-0.7 to -0.3)	-8.3*± 16.3 (-11.7 to -4.8)	-8.8*± 14.1 (-11.8 to -5.8)
5-point GRC T5-T0	81	-1.4*± 5.9 (-2.7 to -0.1)	-3.5*± 14.7 (-6.7 to -0.3)	-0.3*± 0.8 (-0.5 to -0.2)	-5.5*± 13.8 (-8.7 to -2.5)	-7.7*± 12.9 (-10.5 to -4.8)

Change was measured compared to baseline. Negative values represent improvement for CAT, CCQ and SGRQ.
Minimal change scores were reported as mean ± SD (95%CI).
CAT and CCQ were standardised into a scale from zero to 100 to be comparable with SGRQ.
* Significance of change scores at level P < 0.05 at T1/T2/T3/T4/T5 compared to baseline T0.

Abbreviations: 95%CI, 95% confidence interval; CAT, COPD Assessment Test; CCQ, Clinical COPD Questionnaire; COPD, chronic obstructive pulmonary disease; GRC, global rating of change scale; N, number of patients; SD, standard deviation; SGRQ, St. George's Respiratory Questionnaire; T0, baseline measurement; T1, measurement at discharge; T2, 3-months follow-up; T3, 6-months follow-up; T4, 9-months follow-up; T5, 12-months follow-up.

5.4.4 Tests of significance between MCID estimates

ICC values ranged 0.5-0.7 for CAT and CCQ, and 0.6-0.7 for SGRQ (*Table 5*).

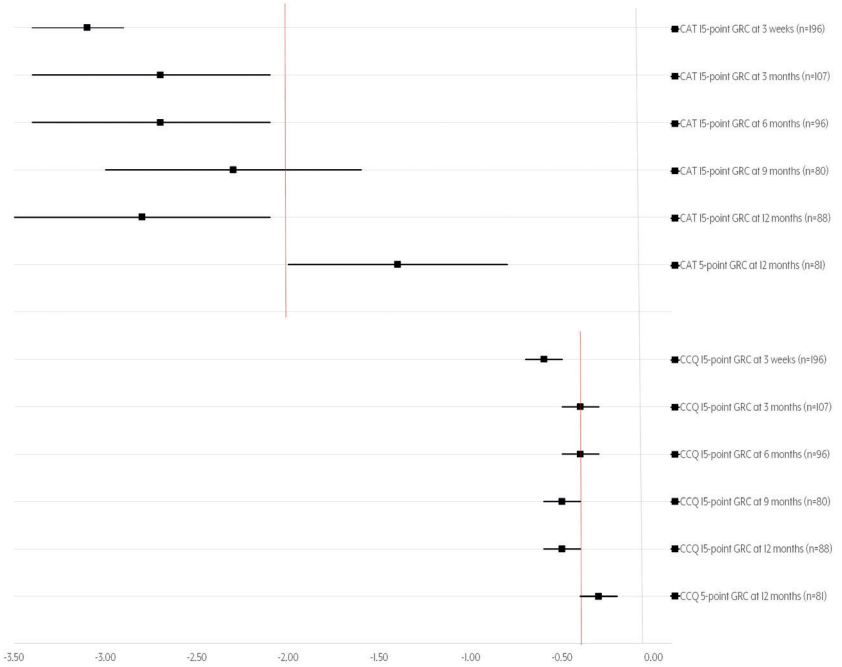
Table 5: Determination of appropriate confidence interval testing for significant different MCIDs

	15-point GRC T0-T1	15-point GRC T0-T2	15-point GRC T0-T3	15-point GRC T0-T4	15-point GRC T0-T5	5-point GRC T0-T5
CAT						
Mean change	-3.1	-2.7	-2.7	-2.3	-2.8	-1.4
SD of the change	5.3	6.4	6.7	6.1	6.7	5.9
N of patients	196	107	96	80	88	81
ICC	0.7	0.6	0.5	0.6	0.5	0.6
Z score required	0.62	0.98	0.98	0.98	0.98	0.98
Standard Error	0.4	0.6	0.7	0.7	0.7	0.7
Adjusted CI	-3.4 to -2.9	-3.4 to -2.1	-3.4 to -2.1	-3.0 to -1.6	-3.5 to -2.1	-2.0 to -0.8
Standardised mean change and adjusted CI	-7.8 (-8.4 to -7.2)	-6.9 (-8.4 to -5.3)	-6.8 (-8.5 to -5.2)	-5.8 (-7.4 to -4.1)	-7.0 (-8.8 to -5.3)	-3.5 (-5.1 to -1.90)
CCQ Total						
Mean change	-0.6	-0.4	-0.4	-0.5	-0.5	-0.3
SD of the change	0.8	1.0	1.1	0.8	1.0	0.8
N of patients	196	107	96	80	88	81
ICC	0.7	0.6	0.5	0.7	0.5	0.6
Z score required	0.62	0.98	0.98	0.62	0.98	0.98
Standard Error	0.1	0.1	0.1	0.1	0.1	0.1
Adjusted CI	-0.7 to -0.5	-0.5 to -0.3	-0.5 to -0.3	-0.6 to -0.4	-0.6 to -0.4	-0.4 to 0.2
Standardised mean change and adjusted CI	-9.3 (-10.0 to -8.7)	-7.3 (-9.0 to -5.7)	-7.0 (-8.8 to -5.2)	-7.7 (-8.7 to -6.7)	-8.3 (-10.0 to -6.7)	-5.5 (-7.0 to -4.0)
SGRQ Total						
Mean change	-8.4	-7.6	-9.2	-10.3	-8.8	-7.7
SD of the change	11.8	13.8	14.0	12.9	14.1	12.9
N of patients	196	107	96	80	88	81
ICC	0.7	0.7	0.7	0.7	0.6	0.6
Z score required	0.62	0.62	0.62	0.62	0.98	0.98
Standard Error	0.8	1.3	1.4	1.4	1.5	1.4
Adjusted CI	-8.9 to -7.9	-8.4 to -6.8	-10.1 to -8.3	-11.2 to -9.4	-10.3 to -7.4	-9.1 to -6.3

Abbreviations: CAT, COPD Assessment Test; CCQ, Clinical COPD Questionnaire; CI, confidence interval; COPD, chronic obstructive pulmonary disease; GRC, global rating of change; ICC, intraclass correlation coefficient; N, number of patients; SD, standard deviation; SGRQ, St. George's Respiratory Questionnaire; T0, baseline measurement; T1, measurement at discharge; T2, 3-months follow-up; T3, 6-months follow-up; T4, 9-months follow-up; T5, 12-months follow-up.

Figures 3 and 4 visually plot the MCID estimates for CAT, CCQ and SGRQ including their respective adjusted CIs for each recall period on both GRCs. Overlap was present for all CAT MCID estimates, except for the 12-months estimate using the 5-point anchor question compared with the 15-point GRC (*Figure 3*). A significantly absolute lower MCID estimate

Figure 3: MCID estimates with for dependency adjusted confidence intervals (CIs) for CAT and CCQ



Data presented as MCID estimates (squares) and respective confidence interval (horizontal line) adjusted for the dependency of the data. The red vertical lines represent the MCID estimates for the CAT and CCQ total score obtained from the literature. Negative values indicate improvement in health status.

Abbreviations: CAT, COPD Assessment Test; CCQ, Clinical COPD Questionnaire; CI, confidence interval; COPD, chronic obstructive pulmonary disease; GRC, global rating of change; N, number of patients; MCID, minimal clinically important difference.

was observed for CAT using the 5-point GRC (*difference -1.4: adjusted CI -2.3 to -0.5*). The MCID measured with the 15-point GRC after 9 months as well as the MCID using the 5-point anchor question overlapped with the CAT estimate from the literature of 2 points.

The MCID plotted for the CCQ visualised that all estimates with their corresponding CIs overlapped one another, except for the estimate after 12 months with the 5-point GRC in comparison to the 15-point GRC (*Figure 3*). A significantly absolute lower MCID estimate was observed for the CCQ using the 5-point GRC at 12 months (*difference -0.2: adjusted CI -0.3 to -0.1*). All estimates included the MCID from the literature of 0.40 points, except for the 15-point GRC anchor question estimate after 3 weeks. The plot for the MCID of the SGRQ showed all ranges overlapping one another, except for the 9-months 15-point GRC anchor question method, which was significantly different from the 15-point GRC estimate after 3 weeks and after 3 months (*Figure 4*). There were no significant differences between the 5-point and 15-point GRC at 12 months. All estimates were significantly different from the 4 points estimate in the literature.

5.5 Discussion

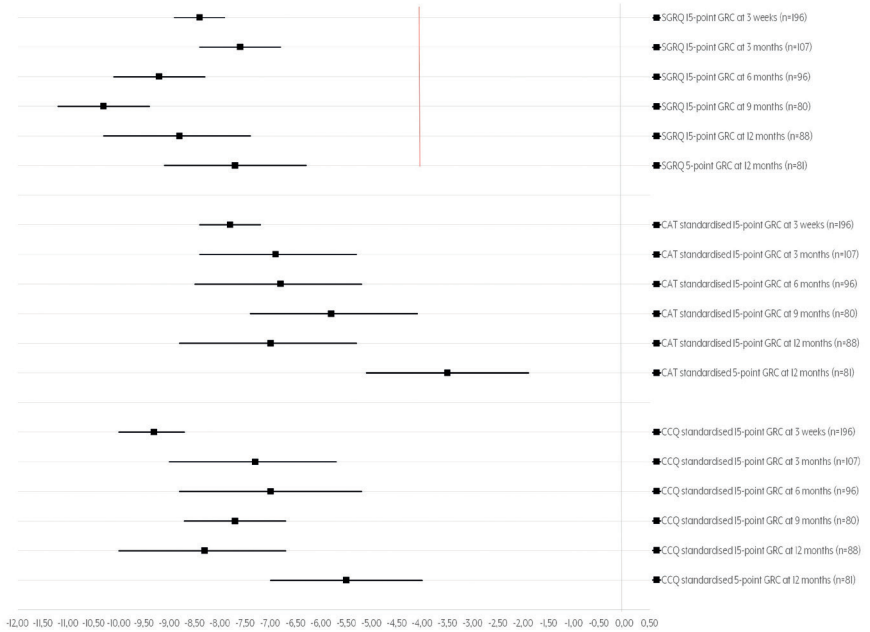
5.5.1 Summary of main findings

This study found no systematic significant differences between various recall periods of a 15-point anchor question on the MCID for improvement of the health status tools CAT, CCQ and SGRQ for patients with COPD in a PR setting. Using this 15-point GRC, MCID estimates for improvement ranged -3.1 to -2.3 for CAT; -0.6 to -0.4 for CCQ; and -10.3 to -7.6 for SGRQ. Higher absolute MCID estimates were observed for CAT and CCQ with a shorter recall period directly after PR, although not significant. The 9-month recall period on the 15-point GRC for the SGRQ was significantly higher in absolute value when comparing with the estimates at 3 weeks and 3 months. However, an anchor question with only 5 answering options did result in significantly absolute lower MCIDs for CAT and CCQ in comparison with the 15-point GRC at 12 months. Estimates were -1.4 for CAT (*significant difference -1.4*), -0.3 for CCQ (*significant difference -0.2*), and -7.7 for the SGRQ (*nonsignificant difference -1.1*).

5.5.2 Interpretation of findings

The MCID ranges found in the current study for both CAT and CCQ were in correspondence with those available in the literature [21-28]. Recall periods on the anchor question of 2-3 days, 3 weeks and 8 weeks have been used before for CAT and CCQ [21, 23, 25, 27-28]. Most MCID estimates for the CAT in the current study were significantly higher than the 2 points threshold, which had been advocated using a 5-point GRC scale [27-28]. Since

Figure 4: MCID estimates with for dependency adjusted confidence intervals (CIs) for SGRQ including standardised estimates for CAT and CCQ



Data presented as MCID estimates (squares) and respective confidence intervals (horizontal line) adjusted for the dependency of the data. The red vertical lines represent the MCID estimates for the CAT and CCQ total score obtained from the literature. Negative values indicate improvement in health status.

Abbreviations: CAT, COPD Assessment Test; CCQ, Clinical COPD Questionnaire; CI, confidence interval; COPD, chronic obstructive pulmonary disease; GRC, global rating of change; N, number of patients; MCID, minimal clinically important difference; SGRQ, St. George's Respiratory Questionnaire.

CAT only allows for integer scores, a cutoff MCID of 3 points would be suggested here. For the CCQ, all recall periods and anchor question types included the 0.40-point MCID as reported in the literature, although our estimates were closer to 0.50 points [21-26]. Both 5- and 15-point GRCs were used generating the original 0.40 MCID estimate for the CCQ [21, 23].

The estimates for the SGRQ in the current study were significantly higher compared with the existing 4 points MCID, which is used extensively in scientific research [29, 31]. This MCID was among others based upon a 5-point question requiring patients with COPD to assess their treatment effects over a 16-week period. It did not require patients to assess their experienced change in health status, hence this may result in a different MCID. The current study provided additional support to the recommendation by Welling *et al.* [30] and Alma *et al.* [25] that the MCID of the SGRQ of 4 points should be set higher.

There was a remarkable significant difference between the 5- and 15-point anchor question GRC scale in estimating the MCIDs for CAT and CCQ at 12 months, although the Spearman correlation between both anchor scales was strong. However, the classification of patients according to both GRCs was only for 55% consistent, resulting thus in a different categorisation of the degree of change assessed by patients themselves. Although the 15-point GRC was analysed as a 7-point scale, the patients had 15 answering options to choose from, compared with just five on the other GRC. Too few reply options on an anchor question might lead to a loss of relevant information, leading to less discriminative power and lower sensitivity [9]. It may result into lower MCIDs. This seems to be the case for the current study for both CAT and CCQ, and to a lesser extent for SGRQ as well. Earlier studies used only 5-point GRCs for CAT and SGRQ [27-29, 31]. These studies showed lower absolute MCIDs. A 5-point anchor scale may therefore not discriminate sufficiently. Kamper SJ *et al.* recommended to include 7-11 reply options for optimal discrimination [9]. Another difference between the current 5- and 15-point GRC was that the first one was a verbal scale, while the latter one was a numeric scale. Possibly this has influenced the classification, as words may result in a different perception in comparison to numbers.

Using an anchor question to determine an instrument's MCID is common practice [6-8]. Jaeschke *et al.* were the first to use this approach in determining the MCID of the Chronic Respiratory Questionnaire (CRQ) using a 15-point Likert scale GRC [5]. Since then many have adopted this method, but have also applied alternative versions to determine the MCID. The approach is easy to administer and the single best measure of the significance of change from the patient's perspective [9-11]. However, anchor questions rely on the patient's ability to recall their former health state [9-11]. Accurate

recall is determined by factors such as forgetting, more recent (impactful) health events, and current mood state [11]. GRC scales may therefore not provide an accurate reflection of the real experienced change due to these recall biases.

It has been speculated that longer recall periods would lead to less accurate estimates of change and even to different MCIDs [10, 38-42]. Evaluation of change turned out to be more correlated with the current health state and severity of experienced symptoms, rather than with the former (baseline) condition [9-10, 41-48]. There are, however, also studies that did not find specific differences between recall periods [39, 49-51]. There is no single optimal recall period [39, 51]. The required window is dependent upon whether or not acute effects need to be measured, whether acute events occur, as well as the nature of the disease [39, 52]. Longer recall periods may therefore be appropriate for chronic conditions with slow changes. It was argued that the optimal length for measuring change on a patient-reported outcome (PRO) in COPD would be six to 12 months [53]. A recall period of more than one year could lead to problems due to the progressive nature of the disease. In addition to the impact of recall bias, a patient's evaluation of a specific health state might change over time due to a response shift [54]. This concept refers to a change in the meaning of the concept HRQoL for the patient. Response shift was demonstrated to have an influence on the MCID for HRQoL tools in breast cancer research [55]. Evidence for the influence of response shift as well as recall bias on the MCID for health status for COPD is currently absent in the literature.

The current study had a fixed recall moment, which was related to the start of an intense PR programme. The effects of PR would be expected to remain over a longer period of time, leading to less exacerbations and less acute changes in the health state of the patients with COPD [56]. Jones *et al.* [53] recommended measurement of PROs in COPD over a 6-12 month period as the optimal recall period, which our study did. The assessment of change compared with the start of PR, the expected stability of COPD symptoms over time after PR and the use of the optimal recall period might help explain why this study found stable MCID estimates during follow-up.

Correlations between the anchor question and the health status change scores were sufficient to be used as anchor, except for the 3-week measurement period. It may, therefore not be surprising, that those estimates were especially for CAT and CCQ higher than the other MCID estimates. Evaluating change directly after an impacting event, such as PR or exacerbations, could potentially bias the MCID measurement of an instrument. The estimates of the SGRQ seemed rather stable over time, perhaps because SGRQ is a more extensive and lengthier tool in comparison to the CAT and CCQ.

5.5.3 Strengths and limitations

This is the first study to investigate the impact of the recall period of the patient's GRC and its design on the MCID for improvement of health status tools for COPD. It is to the best of our knowledge the only study, which measured the MCID of the CAT, CCQ and SGRQ in one study over multiple study periods, and included a unique test of significance for the MCID according to the methods of Afshartous *et al.* [37]. In the current study, MCIDs were tested over multiple periods of time. No correction for multiple testing was made, risking an increase in the probability to run a type I error. However, since this was a diagnostic study, we considered this to be of limited importance as there is no intention to make a formal statement about efficacy or safety based on hypothesis testing [57]. Furthermore, the confidence intervals for the MCID estimates were adjusted for the dependency of multiple follow-up data.

The results found in this study are valid for a PR setting. As MCIDs may differ per setting, the results need not necessarily be valid in other populations [11]. However, our results were in line with the existing MCIDs in the literature, which were also determined outside the field of PR. MCIDs were determined based upon a patient's perspective of their health status change. No clinician, neither the patient, was involved to make a judgement about the clinical relevance of the perceived change though. Correlations between the GRCs and the health status questionnaires were sufficient according to pre-determined criteria, however in fact these correlations are still only small to moderate.

Another limitation is that the data used in this study were based on improvement only, as the number of patients deteriorating for each follow-up period was small to allow for significance testing. MCIDs for improvement may, however, differ from those for deterioration [11]. Furthermore, this study determined the MCID over different recall periods using the 15-point GRC scale. The 5-point anchor question was, however, only measured over a 12-month period. It would not be possible to conclude whether recall bias occurred for the 5-point GRC. Last, the anchor-based MCID technique can be considered a population-based figure, rather than a reflection of the individual's change [6-8, 11]. This is a limitation of the technique in itself. Using a larger sample would lead to regression to the mean of the MCID estimate, which is less subject to larger changes in an individual's health state.

5.5.4 Implications for future clinical practice and theory

No other evidence exists for the impact of the recall period and the design of the anchor question on the determination of MCIDs in health status for COPD. Ideally, more research is needed to confirm or falsify the current findings in both PR and other settings. It would be recommended to use various patient-referencing anchors over multiple periods of

time to carefully estimate an instrument's MCID. Multiple MCIDs might potentially apply to practice for different time periods of measurement used in clinical trials. However, this study was the first to suggest otherwise. It indicated a differentiation might be needed between measurement of change directly after an impacting event and in stable patients, as this may be an important factor influencing recall bias.

5.5.5 Conclusions

Various recall periods on a 15-point anchor question seemed not to be associated with systematic significant differences in the MCIDs for improvement of the CAT, CCQ and SGRQ; with the exception of the shortest 3-week measurement period directly after PR for CAT and CCQ, resulting in absolute higher MCID estimates. Measuring change with a shorter recall period directly after an impacting event might potentially bias measurement. Using an anchor question with less answering options over a one-year period of time in determining an instrument's MCID may also coincide with (significantly) lower absolute MCID estimates as less discriminative options might be available for the patient.

5.6 Declarations

5.6.1 Ethics approval and consent to participate

This study is a secondary analysis of a subsample from the routine inspiratory muscle training within COPD rehabilitation (RIMTCORE) real-life randomised controlled trial (German clinical trial register *#DRKS00004609*) in the Klinik Bad Reichenhall, Center for Rehabilitation, Pulmonology and Orthopedics in Germany. All patients signed informed consent upon participation. The RIMTCORE trial was approved by the Ethik-Kommission der Bayerischen Landesärztekammer (*#12107*).

5.6.2 Funding

The main RIMTCORE trial was funded by the Deutsche Rentenversicherung (German Pension Insurance South Bavaria). The current study regarding the MCID of the CAT, CCQ and SGRQ received financial support from the Junior Scientific Masterclass as part of the University of Groningen.

5.6.3 Acknowledgements

We are grateful to the Junior Scientific Masterclass of the University of Groningen, who financially supported the research position of the first author.

5.6.4 Availability of data and material

The data that support the findings of this study are not publicly available. Participating patients have only agreed upon availability of their data to the Klinik Bad Reichenhall, their scientific partners in the data analysis and the Committee of the Bavarian State Chamber of Labor in Munich.

5.6.5 Authors' contributions and consent

Konrad Schultz, Michael Wittmann, Danijel Jelusic and Michael Schuler planned the RIMTCORE study design regarding the effects of IMT training during pulmonary rehabilitation, and were responsible for data collection. Harma Alma, Corina de Jong, Robbert Sanderman and Thys van der Molen designed the current study regarding the MCID of the CAT, CCQ and SGRQ. Harma Alma, Corina de Jong and Boudewijn Kollen performed the statistical analysis. Harma Alma wrote the first draft, while Corina de Jong, Boudewijn Kollen, Janwillem Kocks, Robbert Sanderman and Thys van der Molen actively participated in the review process. Robbert Sanderman and Thys van der Molen supervised and participated in different steps of the study, as well as in writing. All authors participated in various steps in the study, edited the manuscript and gave their approval for submission.

5.6.7 Competing interests

Harma Alma, Corina de Jong, Danijel Jelusic, Michael Wittmann, Michael Schuler, Boudewijn Kollen and Robbert Sanderman have nothing to disclose. Janwillem Kocks reports personal fees from Novartis; research grants and personal fees from Boehringer Ingelheim; research grants and personal fees from GSK; research grants from Stichting Zorgdraad; personal fees from IPCRG; personal fees from Springer Media; and travel arrangements from Chiesi BV, GlaxoSmithKline BV, and IPCRG, all outside the submitted work. Konrad Schultz received lecture fees from Boehringer Ingelheim, AstraZeneca, Berlin Chemie, Novartis, Chiesi, Mundipharma, Takeda, GSK and MSD, all outside the submitted work. Thys van der Molen reports personal reimbursements from GSK, TEVA, Astra Zeneca, Boehringer Ingelheim and study grants from Astra Zeneca and GSK. After this study was terminated, he became an employee of GSK. None of these stated conflicts of interest are linked to the current manuscript. T. van der Molen developed the CCQ and holds the copyright.

5.7 Supplementary material

5.7.1 15-point global rating of change anchor question

Figure 1: 15-point global rating of change anchor question used during follow-up (in German)

Globale Veränderung des Gesundheitszustandes

Bitte bewerten Sie auf Folgender Skala die Veranderung Ihres **jetzigen Gesundheitszustands gegenüber der Zeit direkt vor Beginn der Rehabilitation**. Bitten kreuzen Sie nur das Kästchen an, das am besten zutrifft:

Im Vergleich zu der Zeit direkt vor Beginn der Rehabilitation ist mein Gesundheitszustand jetzt...

-7	-6	-5	-4	-3	-2	-1	0	+1	+2	+3	+4	+5	+6	+7

Sehr viel schlechter ..... unverändert ..... sehr viel besser

5.7.2 5-point global rating of change anchor question

Figure 2: 5-point global rating of change anchor question used at 12-months follow-up (in German)

	Derzeit viel besser	Derzeit etwas besser	Etwa wie vor einem Jahr	Derzeit etwas schlechter	Derzeit viel schlechter
Im Vergleich zum vervangenen Jahr, wie würden Sie Ihren derzeitigen Gesundheitszustand beschreiben?	1	2	3	4	5

5.8 References

[1] Reardon, JZ, Lareau, SC and ZuWallack, R. Functional Status and Quality of Life in Chronic Obstructive Pulmonary Disease. *The American Journal of Medicine* 2006; 119(10 Suppl 1): S32–S37.

[2] Jones, PW. Health status measurement in chronic obstructive pulmonary disease. *Thorax* 2001; 56(11): 880-887.

[3] Jones, PW. Health Status: What Does It Mean for Payers and Patients? *Proceedings of the American Thoracic Society* 2006; 3(3): 222-226.

[4] Guyatt, GH, Feeny, DH and Patrick, DL. Measuring health-related quality of life. *Annals of Internal Medicine* 1993; 118(8): 622-629.

[5] Jaeschke, R, Singer, J and Guyatt, GH. Measurement of health status. Ascertaining the Minimal Clinically Important Difference. *Controlled Clinical Trials* 1989; 10(4): 407-415.

[6] Copay, AG, Subach, BR, Glassman, SD *et al*. Understanding the minimum clinically important difference: a review of concepts and methods. *The Spine Journal* 2007; 7(5): 541-546.

[7] Revicki, D, Hays, RD, Cella, D *et al*. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *Journal of Clinical Epidemiology* 2008; 61(2): 102-109.

[8] Beaton, DE, Boers, M and Wells, GA. Many faces of the minimal clinically important difference (MCID): a literature review and directions for future research. *Current Opinion in Rheumatology* 2002; 14(2): 109-114.

[9] Kamper, SJ, Maher, C and Mackay, G. Global Rating of Change Scales: A Review of Strengths and Weaknesses and Considerations for Design. *Journal of Manual & Manipulative Therapy* 2009; 17(3): 163-170.

[10] Grovle, L, Haugen, AJ, Hasvik, E *et al*. Patients ratings of global perceived change during 2 years were strongly influenced by the current health status. *Journal of Clinical Epidemiology* 2014; 67(5): 508-515.

[11] Crosby, RD, Kolotkin, RL and Williams, GR. Defining clinically meaningful change in health-related quality of life. *Journal of Clinical Epidemiology* 2003; 56(5): 395-407.

[12] Jones, PW, Price, D and van der Molen, T. Role of clinical questionnaires in optimizing everyday care of chronic obstructive pulmonary disease. *International Journal of COPD* 2011; 6: 289-296.

[13] Curtis, JR and Patrick, DL. The assessment of health status among patients with COPD. *European Respiratory Journal Supplement* 2003; 41: 36s-45s.

[14] Tsiligianni, I, Kocks, J, Tzanakis, N *et al*. Factors that influence disease-specific quality of life or health status in patients with COPD: a review and meta-analysis of Pearson correlations. *Primary Care Respiratory Journal* 2011; 20(3); 257-268.

[15] Westwood, M, Bourbeau, J, Jones, PW *et al*. Relationship between FEV_1 change and patient-reported outcomes in randomized trials of inhaled bronchodilators for stable COPD: a systematic review. *Respiratory Research* 2011; 12: 40.

[16] Jones, PW, Harding, G, Berry, P *et al*. Development and first validation of the COPD Assessment Test. *European Respiratory Journal* 2009; 34(3): 648-654.

[17] Van der Molen, T, Willemse, BWM, Schokker, S *et al*. Development, validity and responsiveness of the Clinical COPD Questionnaire. *Health and Quality of Life Outcomes* 2003, 1: 13.

[18] Jones, PW, Quirck, FH and Baveystock, CM. The St George's Respiratory Questionnaire. *Respiratory Medicine* 1991; 85 (Suppl B): 25-31.

[19] Global initiative for Chronic Obstructive Lung Disease (GOLD). Global Strategy for the Diagnosis, Management and Prevention of COPD (2017 Report). 2017. Available at http://www.goldcopd.org/gold-2017-global-strategy-diagnosis-management-prevention-copd/. Accessed 11th of July 2017.

[20] Kocks, JWH, Blom, CMG, Kasteleyn, MJ *et al*. Feasibility and applicability of the paper and electronic COPD assessment test (CAT) and the clinical COPD questionnaire (CCQ) in primary care: a clinimetric study. *npj Primary Care Respiratory Medicine* 2017; 27(1): 20.

[21] Kocks, JW, Tuinenga, MG, Uil, SM *et al*. Health status measurement in COPD: the minimal clinically important difference of the clinical COPD questionnaire. *Respiratory Research* 2006; 7: 62.

[22] Kon, SS, Dilaver, D, Mittal, M *et al*. The Clinical COPD Questionnaire: response to pulmonary rehabilitation and minimal clinically important difference. *Thorax* 2014; 69(9): 793-798.

[23] Canavan, JL, Dilaver, D, Clark, AL *et al*. Clinical COPD Questionnaire in patients with chronic respiratory disease. *Respirology* 2014; 19(7): 1006-1012.

[24] Tsiligianni, IG, van der Molen, T, Moraitaki, D *et al*. Assessing health status in COPD. A head-to-head comparison between the COPD assessment test (CAT) and the clinical COPD questionnaire (CCQ). *BMC Pulmonary Medicine* 2012, 12: 20.

[25] Alma, H, de Jong, C, Jelusic, D *et al*. T. Health status instruments for patients with COPD in pulmonary rehabilitation: defining a minimal clinically important difference. *npj Primary Care Respiratory Medicine* 2016; 26: 16041.

[26] Smid, DE, Franssen, FM, Houben-Wilke, S *et al*. Responsiveness and MCID Estimates for CAT, CCQ, and HADS in Patients With COPD Undergoing Pulmonary Rehabilitation: A Prospective Analysis. *Journal of the American Medical Directors Association* 2017; 18(1): 53-58.

[27] Kon, SS, Canavan, JL, Jones, SE *et al*. Minimum clinically important difference for the COPD Assessment Test: a prospective analysis. *Lancet Respiratory Medicine* 2014; 2(3): 195-203.

[28] Dodd, JW, Hogg, L, Nolan, J *et al*. The COPD assessment test (CAT): response to pulmonary rehabilitation. A multicenter, prospective study. *Thorax* 2011; 66(5): 425-429.

[29] Jones, PW. St. George's respiratory questionnaire: MCID. *COPD* 2005; 2(1): 75-79.

[30] Welling, JB, Hartman, JE, Ten Hacken, NH *et al*. The minimal important difference for the St George's Respiratory Questionnaire in patients with severe COPD. *European Respiratory Journal* 2015; 46(6): 1598-1604.

[31] Jones, PW and Bosh, TK. Quality of life changes in COPD patients treated with salmeterol. *American Journal of Respiratory Critical Care Medicine* 1997; 155(4): 1283-1289.

[32] Schultz, K, Jelusic, D, Wittmann, M *et al*. Inspiratory muscle training does not improve clinical outcomes in 3-week COPD rehabilitation: results from a randomised controlled trial. *European Respiratory Journal* 2018; 51(1): 1702000.

[33] Gupta, N, Pinto, LM, Morogan, A *et al*. The COPD assessment test: a systematic review. *European Respiratory Journal* 2014; 44(4): 873-884.

[34] Zhou, Z, Zhou, A, Zhao, Y *et al*. Evaluating the Clinical COPD Questionnaire: A systematic review. *Respirology* 2017; 22(2): 251-262.

[35] Ware, JE. SF-36 Health Survey: Manual and Interpretation Guide. The Health Institute, New England Medical Center Boston, Massachusetts.

[36] Terwee, CB, Bot, SD, de Boer, MR *et al*. Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of Clinical Epidemiology* 2007; 60(1): 34-42.

[37] Afshartous, D and Preston, RA. Confidence intervals for dependent data: Equating non-overlap with statistical significance. *Computational Statistics and Data Analysis* 2010; 54(10): 2296-2305.

[38] Guyatt, GH, Norman, GR, Juniper, EF and Griffith, LE. A critical look at transition ratings. *Journal of Clinical Epidemiology* 2002; 55(9): 900-908.

[39] Stull, DE, Klinke Leidy, N, Parasuraman, B *et al*. Optimal recall periods for patient-reported outcomes: challenges and potential solutions. *Current Medical Research Opinion* 2009; 25(4): 929-942.

[40] Kjellson, G, Clarke, P and Gerdtham, UG. Forgetting to remember or remembering to forget: A study of the recall period length in health care survey questions. *Journal of Health Economics* 2014; 35: 34-46.

[41] Schmitt, JS and Abbott, JH. Global ratings of change do not accurately reflect functional change over time in clinical practice. *Journal of Orthopaedics and Sports Physical Therapy* 2015; 45(2): 106-111.

[42] Schmitt, JS and Abbott, JH. Patient global ratings of change did not adequately reflect change over time: a clinical cohort study. *Physical Therapy* 2014; 94(4): 534-542.

[43] Schmitt, J and Di Fabio, RP. The Validity of Prospective and Retrospective Global Change Criterion Measures. *Archives of Physical Medical Rehabilitation* 2005; 86(12): 2270-2276.

[44] Kamper, SJ, Ostelo, RW, Knol, DL *et al*. Global Perceived Effect scales provided reliable assessments of health transition in people with musculoskeletal disorders, but ratings are strongly influenced by current status. *Journal of Clinical Epidemiology* 2010; 63(7): 760-766.

[45] Norman, GR, Stratford, P and Regehr, G. Methodological problems in the retrospective computation of responsiveness to change: the lesson of Cronbach. *Journal of Clinical Epidemiology* 1997; 50(8): 869-879.

[46] Dworkin, RH, Turk, DC, Wyrwich, KW *et al.* Consensus Statement Interpreting the Clinical Importance of Treatment Outcomes in Chronic Pain Clinical Trials: IMMPACT Recommendations. *The Journal of Pain* 2008; 9(2): 105-121.

[47] Herrmann, D. Reporting current, past, and changed health status. What we know about distortion. *Medical Care* 1995; 33(4 Suppl): S89-94.

[48] Meek, PM, Lareau, SC and Anderson, D. Memory for symptoms in COPD patients: how accurate are their reports? *European Respiratory Journal* 2001; 18(3): 474-481.

[49] Seidl, H, Meisinger, C, Kirchberger, I *et al.* Validity of self-reported hospital admissions in clinical trials depends on recall period length and individual characteristics. *Journal of Evaluation in Clinical Practice* 2016; 22(3): 446-454.

[50] Shi, Q, Trask, PC, Wang, XS *et al.* Does recall period have an effect on cancer patients' ratings of the severity of multiple symptoms? *Journal of Pain Symptom Management* 2010; 40(2): 191-199.

[51] Bennett, AV, Amtmann, D, Diehr, P and Patrick, DL. Comparison of 7-day recall and daily diary reports of COPD symptoms and impacts. *Value in Health* 2012; 15(3): 466-474.

[52] Norquist, JM, Girman, C, Fehnel, S *et al.* Choice of recall period for patient-reported outcome (PRO) measures: criteria for consideration. *Quality of Life Research* 2012; 21(6): 1013-1020.

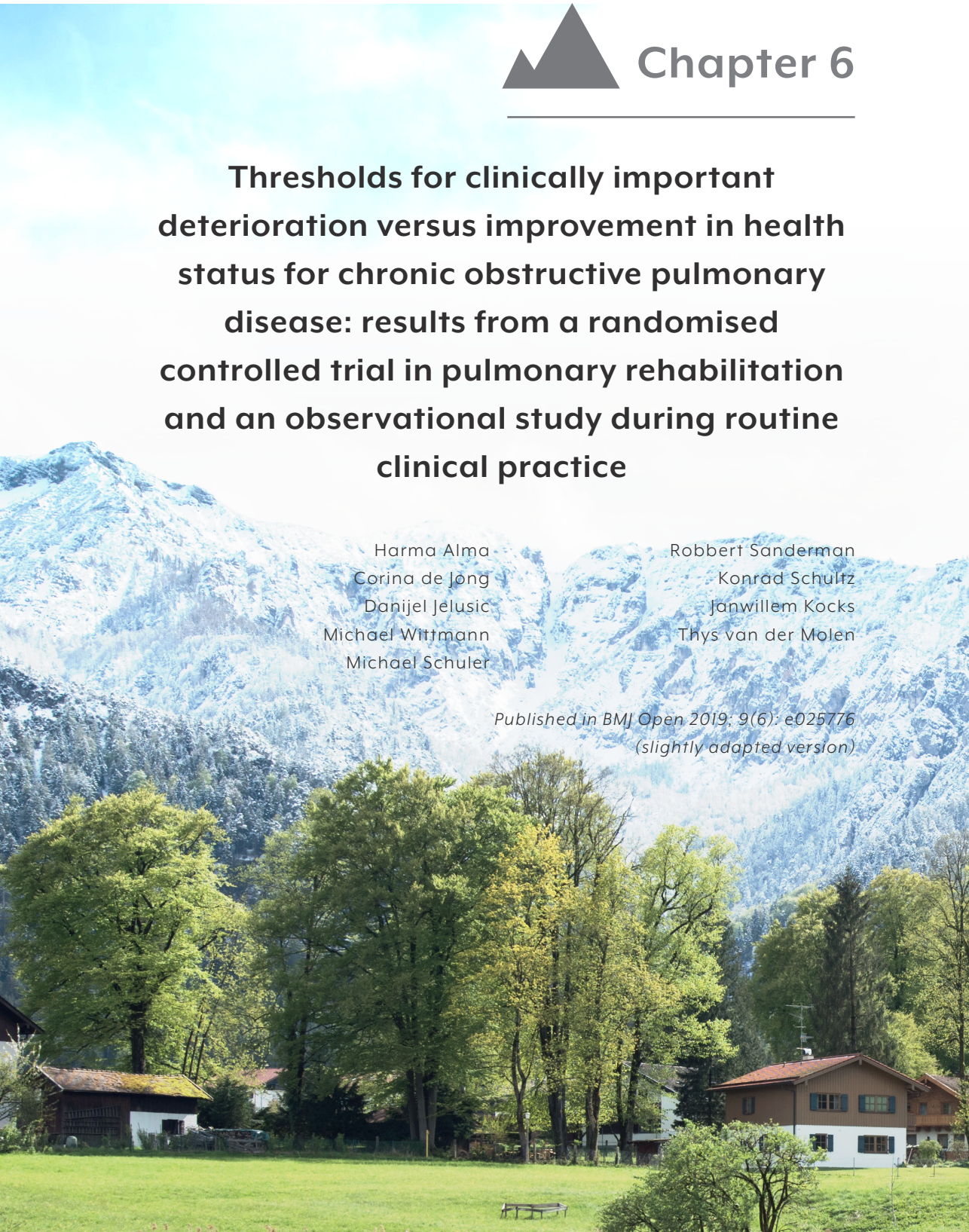
[53] Jones, PW, Beeh, KM, Chapman, KR *et al*. Minimal Clinically Important Differences in Pharmacological Trials. *American Journal of Respiratory Critical Care Medicine* 2014; 189(3): 250-255.

[54] Schwartz, CE and Sprangers, MAG. Methodological approaches for assessing response shift in longitudinal health-related quality-of-life research. *Social Science & Medicine* 1999; 48(11): 1531-1548.

[55] Ousmen, A, Conroy, T, Giullemin F *et al.* Impact of the occurrence of a response shift on the determination of the minimal important difference in a health-related quality of life score over time. *Health and Quality of Life Outcomes* 2016; 14(1): 167.

[56] Lacasse, Y, Goldstein, R, Lasserson, TJ *et al*. Pulmonary rehabilitation for chronic obstructive pulmonary disease (Review). *Cochrane Database of Systematic Reviews* 2006; 4: CD003793.

[57] Phillips, A, Fletcher, C, Atkinson, G *et al.* Multiplicity: discussion points from the Statisticians in the Pharmaceutical Industry multiplicity expert group. *Pharmaceutical Statistics* 2013; 12(5): 255-259.



Chapter 6

Thresholds for clinically important deterioration versus improvement in health status for chronic obstructive pulmonary disease: results from a randomised controlled trial in pulmonary rehabilitation and an observational study during routine clinical practice

Harma Alma
Corina de Jong
Danijel Jelusic
Michael Wittmann
Michael Schuler
Robbert Sanderman
Konrad Schultz
Janwillem Kocks
Thys van der Molen

Published in BMJ Open 2019; 9(6): e025776
(slightly adapted version)

6.1 Abstract

6.1.1 Background

Chronic obstructive pulmonary disease (COPD) is a progressive disease. Preventing deterioration of health status is therefore an important therapy goal. (Minimal) clinically important differences ((M)CIDs) are used to interpret changes observed. It remains unclear whether (M)CIDs are similar for both deterioration and improvement in health status. This study investigates and compares these clinical thresholds for three widely-used questionnaires.

6.1.2 Methods

Data were retrospectively analysed from an in-house 3-week pulmonary rehabilitation (PR) randomised controlled trial in the German Klinik Bad Reichenhall (study 1), and observational research in Dutch primary and secondary routine clinical practice (RCP) (study 2). Patients with COPD aged ≥18 years (study 1) and aged ≥40 years (study 2) without respiratory comorbidities were included for analysis. The COPD Assessment Test (CAT), Clinical COPD Questionnaire (CCQ) and St. George's Respiratory Questionnaire (SGRQ) were completed at baseline, and at 3, 6, and 12 months. A global rating of change (GRC) scale was added at follow-up. Anchor- and distribution-based methods were used to determine clinically relevant thresholds.

6.1.3 Results

In total, 451 patients were included from PR and 207 from RCP. MCIDs for deterioration ranged from 1.30 to 4.21 (CAT), from 0.19 to 0.66 (CCQ), and from 2.75 to 7.53 (SGRQ). MCIDs for improvement ranged from -3.78 to -1.53 (CAT), from -0.50 to -0.19 (CCQ), and from -9.20 to -2.76 (SGRQ). Thresholds for moderate improvement versus deterioration ranged from -5.02 to -3.29 versus 3.89 to 8.14 (CAT), from -0.90 to -0.72 versus 0.42 to 1.23 (CCQ), and from -15.85 to -13.63 versus 7.46 to 9.30 (SGRQ).

6.1.4 Discussion and conclusions

MCID ranges for improvement and deterioration on the CAT, CCQ and SGRQ were somewhat similar. However, estimates for moderate and large change varied and were inconsistent. Thresholds differed between study settings.

6.2 Background

The use of health status questionnaires is recommended by the global initiative for chronic obstructive lung disease (GOLD) for the assessment, evaluation and management of patients with chronic obstructive pulmonary disease (COPD) [1]. The COPD Assessment Test (CAT) [2], the Clinical COPD Questionnaire (CCQ) [3], and the St. George's Respiratory Questionnaire (SGRQ) [4] are frequently used patient-reported health status tools important for clinical practice and scientific research [5], especially since the burden of COPD is high worldwide [6-7].

Various studies have examined clinically relevant thresholds for change on the CAT, CCQ and SGRQ in order to be able to evaluate and interpret treatment effects [8-18]. The minimal clinically important difference (MCID) is a parameter that quantifies this threshold. It has been defined as "*the smallest difference in score, which patients perceive as beneficial and which would mandate a change in the patient's management*" [19]. MCIDs are particularly interesting for health status questionnaires, where a change in its score is not intuitively meaningful. Change exceeding the level of the MCID can be considered clinically relevant, thus justifying therapy and help developing guidelines. It is pivotal that clinically relevant thresholds for change on a health status tool are rigorously studied and analysed carefully.

Most clinical studies that determine the MCID of patient-reported outcomes (PROs) are executed in the context of an intervention such as pharmacotherapy or pulmonary rehabilitation (PR). This usually results in an improvement in the patients' health-related quality of life (HRQoL). MCIDs for improvement have thus been investigated; however, there is a lack of evidence for the MCIDs for deterioration [20]. It remains unclear and debated upon to what extent clinically relevant thresholds for improvement should be similar to those for deterioration [21-24]. Certain studies outside the field of COPD have analysed the MCIDs of PROs and found evidence that values for improvement differed from deterioration [25-29]. On the other hand, there is also evidence that thresholds might be similar [30]. Interpreting worsening of HRQoL is of major importance, since one needs to differentiate between real worsening of patients' status and random variations. Furthermore, the effects of therapy may also halt further deterioration especially for a progressive chronic disease like COPD. So, no relevant worsening or a reduction in clinically relevant deterioration over time might also be considered a success of therapy and in clinical trials [31].

In health status for patients with COPD, the estimated MCID for the CAT score is 2.00 to 3.00 units [11-15, 20], for the CCQ score 0.40 to 0.50 units [8-13, 20], and for the SGRQ score 4.00 to 8.00 units [12, 16-18, 20]. This is valid for improvement only, as there were too few patients with deterioration to investigate. There are currently no studies that specifically investigate clinically relevant thresholds for deterioration on these PROs. It is, however, worrying that up to date, multiple studies included the MCIDs for improvement on these health status instruments for COPD to interpret deterioration observed in clinical trials [32-34]. This study therefore aims to determine and compare clinically relevant thresholds for deterioration and improvement on the health status questionnaires CAT, CCQ and SGRQ in patients with COPD in both a PR and routine clinical practice (RCP) setting.

6.3 Methods

6.3.1 Study subjects

This study was a retrospective analysis of data obtained from two prospective clinical trials. Study 1 was a secondary analysis of a subsample from the routine inspiratory muscle training within COPD rehabilitation (RIMTCORE) real-life randomised controlled trial in the Klinik Bad Reichenhall, Center for Rehabilitation, Pulmonology and Orthopedics in Germany [12, 35]. Patients were recruited on arrival in the clinic between February 2013 and July 2014. Participants were included if they had COPD GOLD grades II-IV, were aged ≥18 years and gave informed consent [12, 35]. Exclusion criteria were the presence of other respiratory comorbidities (e.g. bronchiectasis, asthma, history of bronchial carcinoma, sarcoidosis, tuberculosis), or alpha-1-antitrypsin deficiency.

Study 2 (MCID study) was an observational trial of patients with COPD GOLD grades I-IV aged ≥40 years without other respiratory comorbidities or alpha-1-antitrypsin deficiency. Patients were recruited from Dutch primary and secondary RCP between September 2015 and September 2016. Patients were approached via multiple general practices, hospitals and the Dutch patient lung federation. The study was evaluated by the Medical Ethical Committee of the University Medical Center Groningen (UMCG), the Netherlands. All patients provided written informed consent.

6.3.2 Patient and public involvement

In both studies, patients and the public have not actively been involved during the design of the study, nor in the assessment of the burden. Summary results are disseminated to participating patients after completion.

6.3.3 Study design and data collection

Patients in study 1 participated in an intensive 3-week full-day inpatient PR programme tailored to the patient's individual needs. Details have been presented previously [12, 35]. Patient descriptives and post-bronchodilator spirometry were collected at baseline and discharge in the clinic. Patients in study 2 received routine care from their physician according to national treatment guidelines. Evaluation of health status over a 12-month period was the primary measurement outcome. Patient descriptives and spirometry data were obtained at baseline. Spirometry results were obtained via the including physician after approval of the participant.

The primary outcomes selected from both prospective studies for this retrospective analysis were the CAT (no recall period), CCQ (weekly version) and SGRQ (monthly version). In study 1, these questionnaires were collected at baseline, at PR discharge and during follow-up at 3, 6, 9 and 12 months. Baseline and discharge measurements were taken in the clinic, where patients were blinded to their baseline scores. Follow-up questionnaires were sent by mail. In study 2, all questionnaires were sent by mail and scored at home at baseline, and at 3, 6 and 12 months. For this retrospective analysis baseline and follow-up scores at 3, 6 and 12 months were included, to allow for sufficient time for deterioration in HRQoL, to include various time periods of measurement, and to allow for comparison between both study settings.

The CAT is an 8-item unidimensional scale with item scores ranging 0-5 (0: no impairment; 5: maximum impairment) and a total score summing up to a maximum of 40 [2]. The CCQ consists of 10 items scoring 0-6 (0: no impairment; 6: maximum impairment) [3]. The items cover the domains symptoms (4 items), functional status (4 items) and mental status (2 items). Total and domain scores on the CCQ derive from adding up relevant item scores and dividing this by the number of items. The SGRQ has 50 items classified into the domains symptoms (8 items), activities (16 items) and impact (26 items) [4]. Domain and total SGRQ scores can range from 0-100 (0: no impairment; 100: maximum impairment). A 15-point Likert scale anchor question (global rating of change GRC) was scored retrospectively by the patient at each follow-up visit in both datasets. The GRC required patients to assess their health status for COPD compared with baseline. The answers were marked on a scale from -7 to +7, ranging from very much worse to very much better and zero equaling no change [36-37].

6.3.4 Study methods

All change scores for the total scores of the CAT, CCQ and SGRQ were calculated as the difference between baseline and the respective follow-up visit (3, 6 and 12 months). Negative change on all questionnaires represented improvement, and positive change

deterioration. First, in the anchor-based approach, changes on the health status instruments were classified using the corresponding score on the GRC. Scores of 0 and ±1 on the GRC indicated no change; scores of ±2 and ±3 represented a minimal improvement/deterioration; scores of ±4 and ±5 were summarised as a moderate improvement/deterioration; and scores of ±6 and ±7 indicated a large improvement/deterioration [36-37]. MCID estimates for both improvement and deterioration on the CAT, CCQ and SGRQ were calculated as the mean change scores including 95% confidence interval (95%CI) of those patients indicating a minimal improvement/deterioration (±2 and ±3) on the GRC for each follow-up visit, verifying normality of distribution. Mean estimates including 95%CI were determined in a similar way for patients indicating no change (GRC 0 and ±1), moderate change (GRC ±4 and ±5) and large change (GRC ±6 and ±7). Second, the distribution-based method half standard deviation (0.5SD) of the change score was calculated for improved and deteriorated health status patients at respective follow-up visits [38].

6.3.5 Data analysis

Data analysis was performed using SPSS 24.0 (IBM, Chicago, USA). Descriptives were evaluated at baseline for either frequencies with percentages (%), mean with standard deviation (SD) or median with range. This was depending on the variable characteristics and/or normality of distribution. Health status data on the CCQ, CAT and SGRQ were evaluated at baseline (T0), 3 months (T2), 6 months (T3) and after 12 months (T5). Normality of distribution was verified using skewness and kurtosis. Values between -1 and +1 were considered indicative for normality. Data were checked for floor- and ceiling effects defined as over 15% of patients scoring in the lowest and highest 10% of the maximum scale range [39]. Mean and SD (or median and range) were calculated at each measurement moment for all patients, as well as specifically for patients with improved and deteriorated health status scores. Baseline scores were compared between improving and deteriorating patients, and tested using independent t-tests after verifying normality of distribution. Baseline scores were compared between both datasets (PR vs. RCP) using independent t-tests, Mann-Whitney U tests or $\chi 2$ tests depending on the variable characteristic and/or normality of distribution. Health status change scores were all calculated in comparison with baseline. Follow-up scores were compared with baseline to test for significance of change using paired t-tests verifying normality of distribution.

In order to determine clinically relevant thresholds for change, first correlations between the GRC and the CCQ, CAT and SGRQ respectively were assessed using Pearson or Spearman correlation coefficients depending on normality of distribution. Correlations needed to be ≥0.30 (preferably ≥0.50) to be eligible as anchor [22]. Correlations were

assessed between GRC and questionnaire change scores, and between GRC, baseline and follow-up questionnaire scores to assess for a possible response shift. Next, participants were categorised according to their GRC score at each follow-up. Mean changes (95%CI) for each respective category were determined to define thresholds for clinically relevant change. Significance of change for each GRC class at the respective follow-up visit was compared with baseline and assessed with paired t-tests verifying normality of the data. Last, the 0.5SD of the change score was determined for patients with improved and deteriorating health status change scores separately at each follow-up. Thresholds were compared between both study settings (PR versus RCP).

An absolute overall weighted mean MCID estimate for both improvement and deterioration was calculated at the end by multiplying the number of observations (n) at each follow-up visit times the MCID estimate for that period. The sum was divided by the total number of observations. Anchor- and distribution-based approaches had similar weights. Estimates for improvement and deterioration were compared visually in a plot.

6.4 Results

6.4.1 Patient characteristics

Study 1 included 451 patients with completed baseline data *(Table 1)* [12, 35]. During follow-up 355 patients (78.7%) had completed data at T2, 319 patients (70.7%) at T3, and 309 patients (68.5%) at T5. During the 12-months follow-up 8 patients passed away, 41 dropped out at own request, and a varying number of non-response was present. Study 2 included 207 patients with full baseline data *(Table 1)*, of whom 201 (97.1%) completed T2, 186 (89.9%) T3 and 177 (85.6%) T5. Four patients died, 12 patients discontinued at own request, and a various number of non-response was present. There were no significant baseline differences between completers and non-completers of the 12-months follow-up in both studies, except that significantly more women (28.4%) compared with men (10.0%) did not complete the follow-up during RCP. Significant differences in age, forced expiratory volume in one second percentage predicted (FEV_1%pred) and health status were observed between both studies *(Table 1)*.

6.4.2 Health status scores for improvement and deterioration

In study 1 and 2, CAT, CCQ and SGRQ total were normally distributed at baseline and follow-up. Completed pairs of change scores (follow-up vs. baseline) were included (i.e., pair-wise deletion). Floor- and ceiling effects were negligible. Mean health status baseline scores were significantly different for PR and RCP *(Table 1)*. Overall, 58-59% of patients had improved health status scores (negative change) at T5 after PR, compared with

44-47% during RCP *(Table 2)*. After PR mean changes observed on the CAT questionnaire at T5 were -5.45±4.66 for improvers and 5.47±4.22 for patients, who deteriorated; on the CCQ questionnaire -0.87±0.72 for improvement and 0.83±0.62 for deterioration; and on the SGRQ questionnaire -13.83±10.43 for improvers and 10.19±8.94 for patients, who deteriorated *(Table 2)*. In RCP, these estimates were for the CAT -4.53±3.15 for improvement and 3.88±2.59 for deterioration; for the CCQ -0.54±0.54 for improvement and 0.51±0.39 for deterioration; and for the SGRQ -7.74±9.51 for improvement and 8.46±7.06 for deterioration *(Table 2)*.

Table 1: Baseline patient characteristics

Variable	Study 1: PR	Study 2: RCP	Significance
N	451	207	-
Age (years)[a]	57.87 ± 6.56	66.69 ± 7.91	< 0.001*
Gender (male)[b]	293 (65.0)	121 (58.5)	0.507
FEV_1%pred[a]	50.40 ± 15.11	57.06 ± 21.96	0.001*
GOLD I[b]	-	35 (17.4)	
GOLD II[b]	227 (50.3)	80 (39.8)	
GOLD III[b]	176 (39.0)	61 (30.3)	0.199
GOLD IV[b]	48 (10.6)	25 (12.4)	
Smoking pack years[a]	40 (30-50)	37.5 (22.50-51.25)	0.081
CAT Total[a]	20.23 ± 7.33	18.32 ± 7.22	0.002*
CCQ Total[a]	2.86 ± 1.17	2.12 ± 1.02	< 0.001*
CCQ Symptoms[a]	2.87 ± 1.24	2.48 ± 1.03	< 0.001*
CCQ Functional Status[a]	2.86 ± 1.34	2.28 ± 1.40	< 0.001*
CCQ Mental Status[a]	2.86 ± 1.74	1 (0-1.50)	< 0.001*
SGRQ Total[a]	50.69 ± 17.33	42.88 ± 19.16	< 0.001*
SGRQ Symptoms[a]	63.66 ± 21.77	48.04 ± 24.16	< 0.001*
SGRQ Activities[a]	63.58 ± 19.82	61.48 ± 21.10	0.259
SGRQ Impact[a]	39.21 ± 18.81	30.52 ± 19.73	< 0.001*
mMRC[a]	2 (2-4)	1 (1-2)	< 0.001*

[a] Data expressed as mean ± SD or median (IQR).
[b] Data expressed as frequencies (% of total).
* Significance testing at level P < 0.05 using independent t-tests, Mann-Whitney-U tests or *χ2 tests.*

Abbreviations: CAT, COPD Assessment Test; CCQ, Clinical COPD Questionnaire; COPD, chronic obstructive pulmonary disease; FEV_1%pred, forced expiratory volume in one second % predicted; GOLD, global initiative for chronic obstructive lung disease; mMRC, modified Medical Research Council Dyspnoea scale; N, number of patients; PR, pulmonary rehabilitation; RCP, routine clinical practice; SGRQ, St. George's Respiratory Questionnaire.

There were no baseline differences in terms of age, gender and GOLD classification between patients with improved health status and patients, who deteriorated at T5 in both studies. Patients with a worse (read higher) CAT, CCQ or SGRQ baseline score prior to PR had significantly more improved health status after one year. Patients, who improved during RCP, had a significantly better baseline FEV_1%pred.

Table 2: Health status baseline and change scores for all, improved and deteriorated patients

Patients	Change after 3 months (T2)	N (%)	Change after 6 months (T3)	N (%)	Change after 12 months (T5)	N (%)
CAT						
All patients PR	-1.44* (-2.16 to -0.71)	354	-0.91* (-1.66 to -0.16)	319	-0.89* (-1.68 to -0.11)	309
Improvement PR	-5.45 ± 4.57	227 (64.1)	-5.49 ± 4.33	184 (57.7)	-5.45 ± 4.66	180 (58.3)
Deterioration PR	5.75 ± 4.20	127 (35.9)	5.33 ± 4.10	135 (42.3)	5.47 ± 4.22	129 (41.7)
All patients RCP	0.30 (-0.42 to +1.02)	201	0.18 (-0.53 to +0.90)	186	0.14 (-0.59 to +0.87)	177
Improvement RCP	-4.04 ± 3.33	102 (50.7)	-4.64 ± 3.05	81 (43.5)	-4.53 ± 3.15	79 (44.6)
Deterioration RCP	4.23 ± 3.66	83 (41.3)	3.76 ± 2.88	91 (48.9)	3.88 ± 2.59	86 (48.6)
No change RCP	-	16 (8.0)	-	14 (7.5)	-	12 (6.8)
CCQ Total						
All patients PR	-0.26* (-0.37 to -0.15)	355	-0.11 (-0.23 to +0.01)	319	-0.16* (-0.28 to -0.04)	309
Improvement PR	-0.88 ± 0.71	225 (63.4)	-0.84 ± 0.68	181 (56.7)	-0.87 ± 0.72	180 (58.3)
Deterioration PR	0.82 ± 0.68	130 (36.6)	0.84 ± 0.67	138 (43.3)	0.83 ± 0.62	129 (41.7)
All patients RCP	0.00 (-0.09 to +0.08)	200	0.00 (-0.10 to +0.10)	185	-0.02 (-0.12 to +0.09)	174
Improvement RCP	-0.45 ± 0.37	96 (48.0)	-0.52 ± 0.51	87 (47.0)	-0.54 ± 0.54	77 (44.3)
Deterioration RCP	0.50 ± 0.38	89 (44.5)	0.56 ± 0.46	80 (43.2)	0.51 ± 0.39	88 (50.6)
No change RCP	-	15 (7.5)	-	18 (9.7)	-	9 (5.2)
SGRQ Total						
All patients PR	-5.35* (-6.92 to -3.78)	350	-4.85* (-6.47 to -3.23)	312	-3.94* (-5.67 to -2.21)	306
Improvement PR	-13.11 ± 9.65	237 (67.7)	-13.51 ± 9.88	193 (61.9)	-13.83 ± 10.43	180 (58.8)
Deterioration PR	10.93 ± 10.18	113 (32.3)	8.19 ± 8.92	119 (38.1)	10.19 ± 8.94	126 (41.2)
All patients RCP	-0.52 (-1.77 to +0.73)	198	-1.34 (-2.76 to +0.07)	184	-0.87 (-2.60 to +0.86)	174
Improvement RCP	-6.61 ± 5.58	97 (49.0)	-7.91 ± 5.52	75 (40.8)	-7.74 ± 9.51	81 (46.6)
Deterioration RCP	7.36 ± 5.49	101 (51.0)	7.78 ± 6.18	108 (58.7)	8.46 ± 7.06	92 (52.9)
No change RCP	-	-	-	1 (0.5)	-	1 (0.6)

Change calculated compared with baseline. Negative change representing improvement for CAT, CCQ and SGRQ.
Change scores for all patients reported as mean (95%CI).
Change scores for improvement and deterioration presented as mean ± SD.
*Paired t-tests significant at level $P < 0.05$ testing follow-up versus baseline measurements.

Abbreviations: 95%CI, 95% confidence interval; CAT, COPD Assessment Test; CCQ, Clinical COPD Questionnaire; COPD, chronic obstructive pulmonary disease; N, number of patients; PR, pulmonary rehabilitation; RCP, routine clinical practice; SD, standard deviation; SGRQ, St. George's Respiratory Questionnaire; T2, 3-months follow-up; T3, 6-months follow-up; T5, 12-months follow-up.

6.4.3 Clinically important improvement versus deterioration

Significant correlations between the health status change scores and the GRC ranged respectively for study 1 from -0.33 to -0.41 (CAT), from -0.42 to -0.47 (CCQ), and from -0.48 to -0.54 (SGRQ) *(Table 3)*. These ranges were for study 2 respectively from -0.29 to -0.37, from -0.38 to -0.48, and from -0.35 to -0.44. GRC scores had stronger correlations with the respective follow-up health status score compared with baseline and change scores for both studies.

Table 3: Correlations between health status (change) scores and the GRC

Questionnaire	GRC T2-T0		GRC T3-T0		GRC T5-T0	
	PR (N=355)	RCP (N=201)	PR (N=319)	RCP (N=186)	PR (N=309)	RCP (N=177)
CAT Change Score	-0.33	-0.29	-0.40	-0.30	-0.41	-0.37
CAT T0	-0.31	-0.11	-0.25	-0.22	-0.34	-0.22
CAT T2	-0.56*	-0.31	-0.50*	-0.31	-0.50*	-0.33
CAT T3	-	-	-0.55*	-0.40	-0.59*	-0.34
CAT T5	-	-	-	-	-0.64*	-0.48
CCQ Change Score	-0.40	-0.38	-0.44	-0.40	-0.47	-0.48
CCQ T0	-0.26	-0.14	-0.19	-0.22	-0.29	-0.23
CCQ T2	-0.61*	-0.35	-0.52*	-0.26	-0.54*	-0.33
CCQ T3	-	-	-0.56*	-0.43	-0.59*	-0.39
CCQ T5	-	-	-	-	-0.66*	-0.51*
SGRQ Change Score	-0.48	-0.35	-0.51*	-0.33	-0.54*	-0.44
SGRQ T0	-0.28	-0.13	-0.24	-0.20	-0.32	-0.22
SGRQ T2	-0.62*	-0.29	-0.56*	-0.25	-0.58*	-0.28
SGRQ T3	-	-	-0.61*	-0.35	-0.62*	-0.35
SGRQ T5	-	-	-	-	-0.69*	-0.51*

Data reported as Pearson or Spearman correlation coefficients between the health status (change) scores and the GRC.
* Correlations ≥0.50.

Abbreviations: CAT, COPD Assessment Test; CCQ, Clinical COPD Questionnaire; COPD, chronic obstructive pulmonary disease; GRC, global rating of change; N, number of patients; PR, pulmonary rehabilitation; RCP, routine clinical practice; SGRQ, St. George's Respiratory Questionnaire; T0, baseline measurement; T2, 3-months follow-up; T3, 6-months follow-up; T5, 12-months follow-up.

Tables 4-6 and *figures 1-3* present the clinically relevant thresholds for minimal, moderate and large changes on the CAT, CCQ and SGRQ during PR and RCP. On the CAT anchor- and distribution-based estimates ranged from -2.80 to -2.17 *(weighted mean -2.51)* for minimal improvement and from 2.05 to 4.21 for minimal deterioration *(weighted mean 2.76)* during PR *(Table 4, Figure 1)*. These ranges were respectively, from -3.78 to -1.53 *(weighted mean -2.49)* and from 1.30 to 1.97 *(weighted mean 1.65)* during

Figure 1: Forest plot of clinically relevant thresholds for improvement and deterioration on the CAT

Data presented as mean estimates (squares) including 95%CI (horizontal lines). Estimates from the half SD analysis presented as single squares. Weighted mean estimates presented as larger diamonds. Data separated as minor, moderate and large improvement thresholds (left half), versus minor and moderate deterioration thresholds (right half).

Abbreviations: 95%CI, 95% confidence interval; CAT, COPD Assessment Test; COPD, chronic obstructive pulmonary disease; PR, pulmonary rehabilitation; RCP, routine clinical practice; SD, standard deviation; T0, baseline measurement; TT3, 3-months follow-up; T6, 6-months follow-up; T12, 12-months follow-up.

Table 4: Estimates for clinically relevant thresholds for improvement and deterioration on the CAT

CAT	T2-T0		T3-T0		T5-T0		Weighted threshold	
Change	Improvement	Deterioration	Improvement	Deterioration	Improvement	Deterioration	Improvement	Deterioration
Minimal change								
N Anchor-based PR	107	36	96	42	88	43	291	121
Anchor-based PR	-2.74	2.71	-2.73	3.21	-2.80	4.21	-2.75	3.42
N Anchor-based RCP	12	27	14	36	18	46	32	82
Anchor-based RCP	-	-	-2.86	1.97	-3.78	1.63	-3.38	1.78
N Distribution-based PR	227	127	184	135	180	129	591	391
Distribution-based PR	-2.29	2.10	-2.17	2.05	-2.33	2.11	-2.26	2.09
N Distribution-based RCP	102	83	81	91	79	86	262	260
Distribution-based RCP	-1.67	1.83	-1.53	1.44	-1.58	1.30	-1.60	1.52
Moderate change								
N Anchor-based PR	51	9	45	7	37	10	133	17
Anchor-based PR	-5.02	-	-3.29	8.14	-4.27	6.30	-4.23	7.06
N Anchor-based RCP	5	8	12	9	5	9	-	9
Anchor-based RCP	-	-	-	-	-	3.89	-	3.89
Large change								
N Anchor-based PR	16	3	12	2	14	3	42	-
Anchor-based PR	-4.19	-	-7.00	-	-6.07	-	-5.62	-
N Anchor-based RCP	4	3	0	2	9	4	13	4
Anchor-based RCP	-6.00	-	-	-	-4.22	5.75	-4.77	5.75
No change								
N Anchor-based PR	133		115		114		362	
Anchor-based PR	0.03		-0.01		-0.33		-0.10	
N Anchor-based RCP	141		113		83		337	
Anchor-based RCP	-0.16		-0.54		-0.47		-0.36	

Data reported as clinically relevant threshold or N. Negative change represented improvement for all health status instruments.
Paired t-tests applied with significance at level P <0.05.
Non-significant results excluded, except for the "No change" group on the GRC.

Abbreviations: CAT, COPD Assessment Test; COPD, chronic obstructive pulmonary disease; GRC, global rating of change; N, number of patients; PR, pulmonary rehabilitation; RCP, routine clinical practice; T0, baseline measurement; T3, 3-months follow-up; T6, 6-months follow-up; T12, 12-months follow-up.

Table 5: Estimates for clinically relevant thresholds for improvement and deterioration on the CCQ

CCQ	T2-T0		T3-T0		T5-T0		Weighted threshold	
Change	Improvement	Deterioration	Improvement	Deterioration	Improvement	Deterioration	Improvement	Deterioration
Minimal change								
N Anchor-based PR	107	36	96	42	88	43	291	121
Anchor-based PR	-0.44	0.42	-0.42	0.48	-0.50	0.66	-0.45	0.53
N Anchor-based RCP	12	27	14	36	18	46	32	82
Anchor-based RCP	-	-	-0.44	0.46	-0.38	0.33	-0.41	0.39
N Distribution-based PR	225	130	181	138	180	129	586	397
Distribution-based PR	-0.36	0.34	-0.34	0.34	-0.36	0.31	-0.35	0.33
N Distribution-based RCP	96	89	87	80	77	88	260	257
Distribution-based RCP	-0.19	0.19	-0.26	0.23	-0.27	0.20	-0.24	0.21
Moderate change								
N Anchor-based PR	51	9	45	7	37	10	133	7
Anchor-based PR	-0.86	-	-0.72	1.23	-0.90	-	-0.82	1.23
N Anchor-based RCP	5	8	12	9	5	9	-	17
Anchor-based RCP	-	0.85	-	-	-	0.42	-	0.62
Large change								
N Anchor-based PR	16	3	12	2	14	3	42	-
Anchor-based PR	-0.96	-	-1.03	-	-1.18	-	-1.05	-
N Anchor-based RCP	4	3	0	2	9	4	9	4
Anchor-based RCP	-	-	-	-	-1.12	0.98	-1.12	0.98
No change								
N Anchor-based PR	133		115		114		362	
Anchor-based PR	-0.07		0.17		0.10		0.06	
N Anchor-based RCP	141		113		83		337	
Anchor-based RCP	-0.03		-0.10		-0.04		-0.06	

Data reported as clinically relevant threshold or N. Negative change represented improvement for all health status instruments.
Paired t-tests applied with significance at level $P < 0.05$.
Non-significant results excluded, except for the "No change" group on the GRC.

Abbreviations: CCQ, Clinical COPD Questionnaire; COPD, chronic obstructive pulmonary disease; GRC, global rating of change; N, number of patients; PR, pulmonary rehabilitation; RCP, routine clinical practice; T0, baseline measurement; T3, 3-months follow-up; T6, 6-months follow-up; T12, 12-months follow-up.

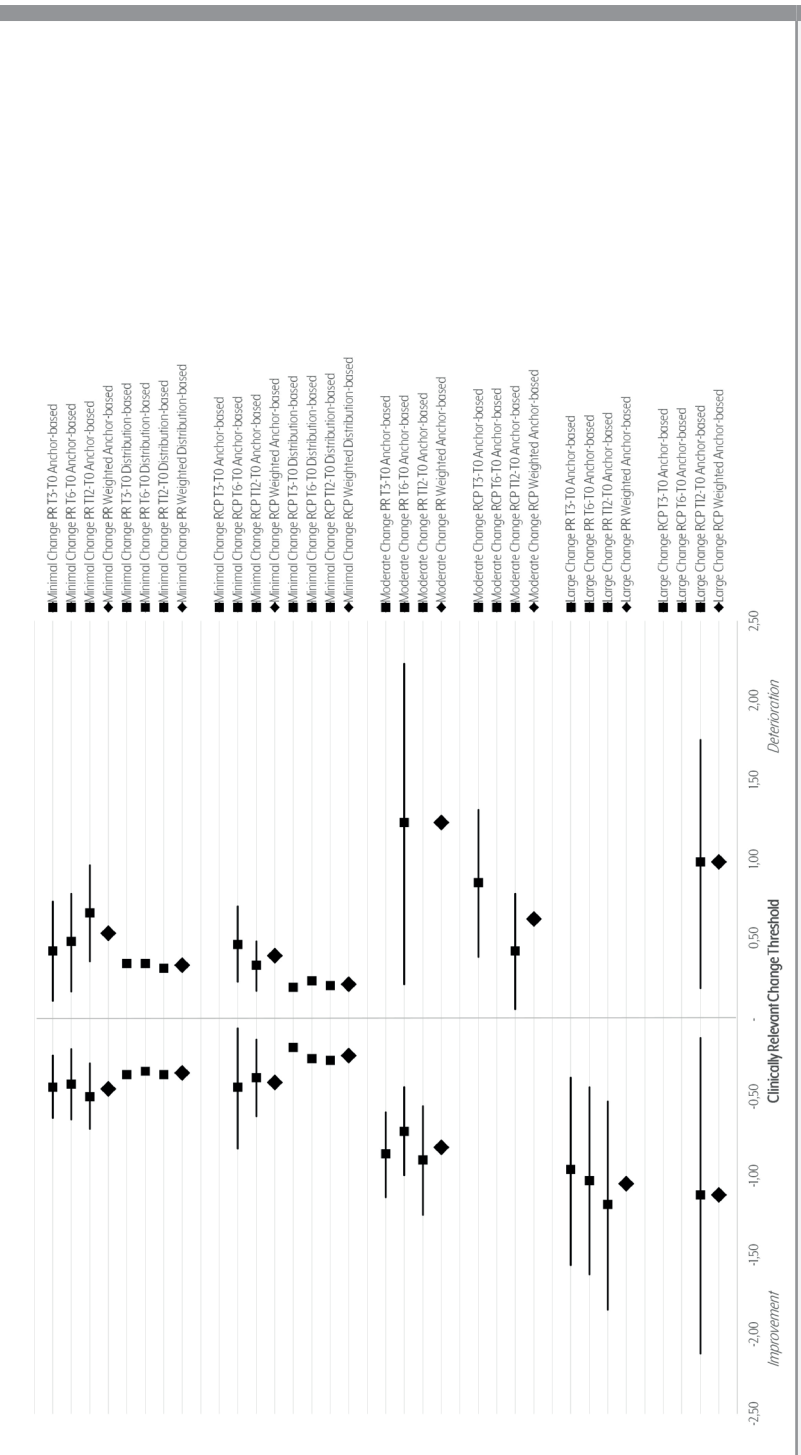
Table 6: Estimates for clinically relevant thresholds for improvement and deterioration on the SGRQ

SGRQ	T2-T0		T3-T0		T5-T0		Weighted threshold	
Change	Improvement	Deterioration	Improvement	Deterioration	Improvement	Deterioration	Improvement	Deterioration
Minimal change								
N Anchor-based PR	107	36	96	42	88	43	291	121
Anchor-based PR	-7.58	5.01	-9.20	5.14	-8.82	7.52	-8.49	5.95
N Anchor-based RCP	12	27	14	36	18	46	14	82
Anchor-based RCP	-	-	-4.70	7.53	-	5.60	-4.70	6.45
N Distribution-based PR	237	113	193	119	180	126	610	358
Distribution-based PR	-4.83	5.09	-4.94	4.46	-5.22	4.47	-4.98	4.66
N Distribution-based RCP	97	101	75	108	81	92	253	301
Distribution-based RCP	-2.79	2.75	-2.76	3.09	-4.76	3.53	-3.41	3.11
Moderate change								
N Anchor-based PR	51	9	45	7	37	10	124	10
Anchor-based PR	-15.85	-	-13.63	-	-15.40	9.30	-16.06	9.30
N Anchor-based RCP	5	8	12	9	5	9	-	9
Anchor-based RCP	-	-	-	-	-	7.46	-	7.46
Large change								
N Anchor-based PR	16	3	12	2	14	3	42	-
Anchor-based PR	-18.33	-	-21.99		-20.58	-	-20.13	-
N Anchor-based RCP	4	3	0	2	9	4	9	-
Anchor-based RCP	-	-	-	-	-18.70	-	-18.70	-
No change								
N Anchor-based PR	133		115		114		362	
Anchor-based PR	-1.50		-0.99		-0.06		-0.88	
N Anchor-based RCP	141		113		83		337	
Anchor-based RCP	0.51		0.19		0.10		0.30	

Data reported as clinically relevant threshold or N. Negative change represented improvement for all health status instruments.
Paired t-tests applied with significance at level $P < 0.05$.
Non-significant results excluded, except for the "No change" group on the GRC.

Abbreviations: GRC, global rating of change; N, number of patients; PR, pulmonary rehabilitation; RCP, routine clinical practice; SGRQ, St. George's Respiratory Questionnaire; T0, baseline measurement; T2, 3-months follow-up; T3, 6-months follow-up; T5, 12-months follow-up.

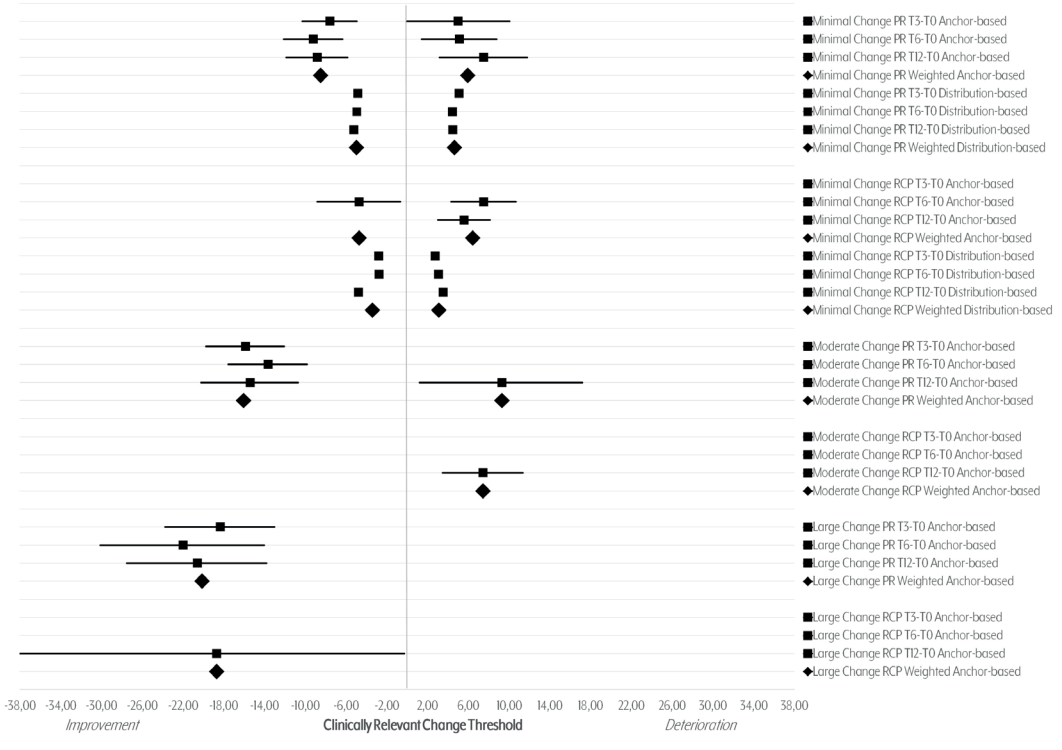
Figure 2: Forest plot of clinically relevant thresholds for improvement and deterioration on the CCQ



Data presented as mean estimates (squares) including 95%CI (horizontal lines). Estimates from the half SD analysis presented as single squares. Weighted mean estimates presented as larger diamonds. Data separated as minor, moderate and large improvement thresholds (left half), versus minor and moderate deterioration thresholds (right half).

Abbreviations: 95%CI, 95% confidence interval; CCQ, Clinical COPD Questionnaire; COPD, chronic obstructive pulmonary disease; PR, pulmonary rehabilitation; RCP, routine clinical practice; SD, standard deviation; T0, baseline measurement; T2, 3-months follow-up; T3, 6-months follow-up; T5, 12-months follow-up.

Figure 3: Forest plot of clinically relevant thresholds for improvement and deterioration on the SGRQ



Data presented as mean estimates (squares) including 95%CI (horizontal lines). Estimates from the half SD analysis presented as single squares. Weighted mean estimates presented as larger diamonds. Data separated as minor, moderate and large improvement thresholds (left half), versus minor and moderate deterioration thresholds (right half).

Abbreviations: 95%CI, 95% confidence interval; PR, pulmonary rehabilitation; RCP, routine clinical practice; SGRQ, St. George's Respiratory Questionnaire; SD, standard deviation; T0, baseline measurement; T2, 3-months follow-up; T3, 6-months follow-up; T5, 12-months follow-up.

RCP. On the CCQ minimal clinically important improvements were determined at -0.50 to -0.34 *(weighted mean -0.40)* for PR and -0.44 to -0.19 *(weighted mean -0.33)* for RCP *(Table 5, Figure 2)*. These thresholds for deterioration were from 0.31 to 0.66 *(weighted mean 0.43)* during PR and from 0.19 to 0.46 *(weighted mean 0.30)* during RCP. On the SGRQ estimates ranged from -9.20 to -4.83 *(weighted mean -6.74)* for minimal improvement and from 4.46 to 7.52 for minimal deterioration *(weighted mean 5.31)* during PR *(Table 6, Figure 3)*. These ranges were respectively, from -4.76 to -2.76 *(weighted mean -4.06)* and from 2.75 to 7.53 *(weighted mean 4.78)* during RCP.

6.5 Discussion

6.5.1 Summary of main findings

Using both anchor- and distribution-based methods, the *weighted* MCIDs for improvement and deterioration on the CAT were, respectively, -2.51 versus 2.76 during PR, and -2.49 versus 1.65 during RCP. These thresholds for improvement and deterioration on the CCQ were, respectively, -0.40 versus 0.43 during PR, and -0.33 versus 0.30 during RCP. MCIDs for the SGRQ were, respectively, -6.74 versus 5.31 during PR, and -4.06 versus 4.78 during RCP for improvement and deterioration. Estimates for minimal clinically important improvement and deterioration were overall somewhat similar; however, absolute MCIDs differed between PR and RCP. Thresholds for moderate and large improvement and deterioration differed from each other, as well as between study settings.

6.5.2 Interpretation of findings

Little evidence exists whether MCIDs for improvement are similar for deterioration [21, 23, 40]. Jaeschke *et al*. were the first to determine the MCID of a health status tool using a 15-point GRC combining both improved and deteriorated patients with COPD into one group of minimally changed participants [19]. Juniper *et al*. elaborated on this by separating minimally improved patients from deterioration in asthma, but only a limited number of patients indicated deterioration and no conclusions on the MCID for deterioration were drawn [37]. Outside the field of COPD, Crosby *et al*. and de Vet *et al*. stated that some studies demonstrated that a smaller MCID for improvement was required compared with deterioration [21, 40]. The current study does not confirm this, although MCIDs seemed smaller for RCP patients compared with PR. Patients experienced more change (hence larger absolute MCIDs) during intervention, possibly as a result of treatment. In RCP, smaller changes may be noted and regarded as relevant for the patient. Up to now it remains unclear, whether the reported differences between PR and RCP are a rehab-specific finding or generally as a result of intervention. Overall, the absolute values of the MCIDs for improvement versus deterioration did not seem to differ much here, with the exception of the SGRQ during PR.

The ranges found in this study for the MCID of the CAT (*improvement* -3.78 to -1.53; *deterioration* 1.30 to 4.21) matched with estimates found in other studies [11-15, 20]. Two studies used a patient-assessed GRC to estimate the MCID of the CAT [14-15]. However, no results were reported for worsened patients or the numbers of patients were too few. Other anchor-based methods suggested that a change of 1 point on the CAT might represent the MCID for deterioration [14]. The weighted thresholds for minimal clinically relevant improvement (-2.51 in *PR* and -2.49 in *RCP*) seemed somewhat comparable with the ones for deterioration (2.76 in *PR* and 1.65 in *RCP*) in the current study, except for deterioration during RCP. As CAT allows only integer scores [2], a change of 3 points seems a valid threshold for improvement and deterioration, although the MCID for deterioration in RCP could be closer to 2 points. Thresholds for moderate improvement (-4.23 in *PR*) and moderate deterioration (7.06 in *PR* and 3.89 in *RCP*) turned out less similar. The number of patients moderately deteriorating was low and differences were observed between both study settings. Moderate change might be experienced with a change on the CAT score of 4-7 points. Two previous studies suggested that a cutoff value of 4 points was identified for acute HRQoL deterioration in clinical practice [41-42]. This would match our estimates for moderate change. The number of patients with a large change was too low with wide CIs to enable valid conclusions.

Regarding the CCQ, the MCID ranges found for both improvement (-0.50 to -0.19) and deterioration (0.19 to 0.66) overlapped each other in absolute sense, indicating that estimates for improvement and deterioration may be similar. However, differences were noted between PR (±0.40) and RCP (±0.30) for both minimal improvement and deterioration. These estimates for the MCID matched with earlier evidence [8-13]. One other study used a GRC to determine the MCID of the CCQ [8]. Unfortunately, no data were available on worsening patients. Thresholds for moderate change on the CCQ were broad (±0.62 to ±1.23). Few patients experienced large changes, but estimates for both types of MCID from both study settings were approximately 1 point.

Minimal thresholds for improvement (-9.20 to -2.76) and deterioration (2.75 to 7.53) on the SGRQ overlapped each other, although more variation was present here. A change of approximately 4-7 points for both improvement and deterioration seemed to be the minimal clinically important threshold in the current study. The MCID for improvement during PR (-6.74) was larger than for deterioration (5.31); however, CIs for deterioration were wide. Estimates for the thresholds during RCP (4-5 points) were smaller compared with PR (5-7 points). Moreover, the distribution-based estimates turned out smaller than the anchor-based estimates, lowering the absolute weighted MCIDs. Thresholds for moderate improvement and deterioration in the current study were not very similar

ranging absolutely from 7.46 to 16.06 points. Estimates for clinically relevant large HRQoL improvement on the SGRQ ranged from -20 to -18 points for PR and RCP, but too few patients were included to draw valid conclusions.

The SGRQ MCID matched to some extent with previous results [12, 16-18, 20]. Jones *et al*. published a threshold of 4 points, which is generally accepted and applied in practice [16, 18]. Interestingly, most results in our current study suggest a larger MCID, although estimates from RCP included this 4 point estimate. The estimate by Jones *et al*. was based on a study using patient preference-based techniques in COPD by applying a 5-point patients' judgement of treatment efficacy of Salmeterol. This MCID of 4 points was valid for the group of patients that experienced effective treatment. In addition, a clinicians' 5-point GRC was scored, resulting in an MCID of 4 points. Clinicians' and patients' ratings are, however, not necessarily similar [43].

6.5.3 Strengths and limitations of the current study

This retrospective analysis of two prospective studies was the first to investigate clinically relevant thresholds for minimal, moderate and large changes in COPD health status comparing both improvement and deterioration using a combination of both anchor- and distribution-based methods. There were sufficient correlations between the GRC and respective health status questionnaires as required [22], although they were still only weak to moderate. It should be noted that correlations were stronger with the follow-up score compared with the baseline and/or change score, possibly due to a response shift. Another strength is that multiple follow-up visits were included to limit possible influence of the period of measurements on the MCID and recall bias [21, 24]. Moreover, this study investigated clinically relevant thresholds for both PR and RCP, improving its clinical application and external validity.

Although this is the first study to investigate thresholds for clinically relevant deterioration, still a limited number of patients indicated deterioration in HRQoL after PR and during RCP. This is a major limitation lowering the statistical power of the analysis, especially since sample size calculations were not based on the separate GRC categories. A second limitation is that the found thresholds demonstrated broad ranges with wide CIs, limiting its accuracy and requiring a larger sample size than our current studies had. Third, it should be taken into account that anchor- and distribution-based approaches each has its own relevance, either based on clinical retrospective assessments or statistical parameters. It is recommended to combine both methods in measuring an instrument's MCID [22]; however, estimates were rather different between these methods.

6.5.4 Implications for future research and clinical practice

Patients with COPD tend to have worsening HRQoL over time; hence, MCIDs for deterioration have an important implication for clinical practice [44-45]. Clinicians and researchers should be able to judge whether groups of patients were really worsening over time or that change observed was subject to random fluctuation. Preventing clinically relevant deterioration in HRQoL by means of therapy is thus an important goal too. Ideally, more research is needed to validate our thresholds for clinically relevant deterioration on the CAT, CCQ and SGRQ for instance in studies of other kinds of interventions than PR. One cannot directly transform the thresholds for improvement into those for deterioration. Evidence outside the field of COPD has found differences. However, in the current study, the estimates turned out rather similar with differing MCIDs between studies. Setting could thus potentially impact the MCID, implying that the results in the current study do not necessarily need to be valid in other settings too.

6.5.5 Conclusions

Determining deterioration in HRQoL is of importance, since one needs to differentiate between real worsening of patients' status and random variations. In this study, estimates for clinically relevant thresholds for improvement and deterioration were somewhat similar, but differed between PR and RCP. It could be recommended to use the following cutpoints for minimal important changes: CAT ≥3 (intervention), CAT ≥2 (RCP), CCQ ≥0.40 (intervention), CCQ ≥0.30 (RCP), SGRQ ≥6 (intervention) and SGRQ ≥5 (RCP) for both minimal improvement and deterioration. Thresholds for respectively moderate and large changes should be further explored, but could approximately be in the range of respectively 4-5 and 5-6 for CAT, 0.80 and 1.00 for CCQ, and 10-15 points and 15-20 points for SGRQ.

6.6 Declarations

6.6.1 Ethics approval and consent to participate

This study is a secondary retrospective analysis of a subsample from the routine inspiratory muscle training within COPD rehabilitation (RIMTCORE) real-life randomised controlled trial (German clinical trial register #DRKS00004609) in the Klinik Bad Reichenhall, Center for Rehabilitation, Pulmonology and Orthopedics in Germany; and a primary analysis of all patients participating in the Dutch observational trial (MCID study) on health status in patients with COPD in routine clinical practice (UMCG trial #201500447). All patients in both studies signed informed consent upon participation. The RIMTCORE trial has been approved by the Ethik-Kommission der Bayerischen Landesärztekammer (#12107). The MCID study has been evaluated by its Medical Ethical Committee.

6.6.2 Funding

The RIMTCORE trial with patients in pulmonary rehabilitation was funded by the Deutsche Rentenversicherung. The Dutch observational study on health status in patients with COPD in routine clinical practice, as well as the current combined retrospective analysis of both prospective studies, received financial support from the Junior Scientific Masterclass as part of the University of Groningen, the Netherlands.

6.6.3 Acknowledgements

We are grateful to the Junior Scientific Masterclass of the University of Groningen, who financially supported the research position of the first author. We would also like to acknowledge all participating patients in both the RIMTCORE trial and the MCID study.

6.6.4 Availability of data and material

The data that support the findings of this study are not publicly available. Participating patients in the RIMTCORE trial have only agreed on the availability of their data to the Klinik Bad Reichenhall, their scientific partners in the data analysis and the Committee of the Bavarian State Chamber of Labor in Munich. Participating patients in the MCID study only agreed on the availability of their data to the University Medical Center Groningen (UMCG) and their scientific partners in the data analysis.

6.6.5 Authors' contributions and consent

Konrad Schultz, Michael Wittmann, Danijel Jelusic and Michael Schuler planned the RIMTCORE study design and were responsible for data collection. Harma Alma, Corina de Jong, Robbert Sanderman, Janwillem Kocks and Thys van der Molen designed the Dutch observational study on COPD health status in routine clinical practice, as well as the current retrospective analysis of both prospective studies. Harma Alma and Corina de Jong performed the statistical analysis. Harma Alma wrote the first draft, while Corina de Jong, Robbert Sanderman, Janwillem Kocks and Thys van der Molen actively participated in the review process. Robbert Sanderman and Thys van der Molen supervised and participated in the different steps of the study, as well as in writing. All authors participated in various steps in the study, edited the manuscript and gave their approval for submission.

6.6.6 Competing interests

Harma Alma, Corina de Jong, Danijel Jelusic, Michael Wittmann, Michael Schuler and Robbert Sanderman have nothing to disclose. Janwillem Kocks reports personal fees from Novartis; research grants and personal fees from Boehringer Ingelheim; research grants and personal fees from GlaxoSmithKline (GSK); research grants from Stichting Zorgdraad; personal fees from the International Primary Care Respiratory Group

(IPCRG); personal fees from Springer Media; and travel arrangements from Chiesi, GSK and IPCRG, all outside the submitted work. Konrad Schultz received lecture fees from Boehringer Ingelheim, AstraZeneca, Berlin-Chemie, Novartis, Chiesi, Mundipharma, Takeda, GSK and MSD, all outside the submitted work. Thys van der Molen reports personal reimbursements from GSK, TEVA, AstraZeneca and Boehringer Ingelheim, and study grants from AstraZeneca and GSK. After this study was terminated, he became an employee of GSK. None of these stated conflicts of interest are linked to the current manuscript. Thys van der Molen developed the Clinical COPD Questionnaire (CCQ) and holds the copyright.

6.7 References

[1] Global initiative for Chronic Obstructive Lung Disease (GOLD). Global Strategy for the Diagnosis, Management and Prevention of COPD (2017 Report). 2017. Available at http://www.goldcopd.org/gold-2017-global-strategy-diagnosis-management-prevention-copd/. Accessed 17th of July 2017.

[2] Jones, PW, Harding, G, Berry, P *et al.* Development and first validation of the COPD Assessment Test. *European Respiratory Journal* 2009; 34(3): 648-654.

[3] Van der Molen, T, Willemse, BWM, Schokker, S *et al.* Development, validity and responsiveness of the Clinical COPD Questionnaire. *Health and Quality of Life Outcomes* 2003, 1: 13.

[4] Jones, PW, Quirck, FH and Baveystock, CM. The St George's Respiratory Questionnaire. *Respiratory Medicine* 1991; 85 (Suppl B): 25-31.

[5] Cave, J, Atkinson, L, Tsiligianni, IG *et al.* Assessment of COPD wellness tools for use in primary care: an IPCRG initiative. *International Journal of COPD* 2012; 7: 447-456.

[6] Lozano, R, Naghavi, M, Foreman, K *et al.* Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet* 2012; 380(9859): 2095-2128.

[7] Adeloye, D, Chua, S, Lee, C *et al.* Global and regional estimates of COPD prevalence: Systematic review and meta-analysis. *Journal of Global Health* 2015; 5(2): 020415.

[8] Kocks, JW, Tuinenga, MG, Uil, SM *et al.* Health status measurement in COPD: the minimal clinically important difference of the clinical COPD questionnaire. *Respiratory Research* 2006; 7: 62.

[9] Kon, SS, Dilaver, D, Mittal, M *et al.* The Clinical COPD Questionnaire: response to pulmonary rehabilitation and minimal clinically important difference. *Thorax* 2014; 69(9): 793-798.

[10] Canavan, JL, Dilaver, D, Clark, AL *et al.* Clinical COPD Questionnaire in patients with chronic respiratory disease. *Respirology* 2014; 19(7): 1006-1012.

[11] Tsiligianni, IG, van der Molen, T, Moraitaki, D *et al.* Assessing health status in COPD. A head-to-head comparison between the COPD assessment test (CAT) and the clinical COPD questionnaire (CCQ). *BMC Pulmonary Medicine* 2012, 12: 20.

[12] Alma, H, de Jong, C, Jelusic, D *et al.* Health status instruments for patients with COPD in pulmonary rehabilitation: defining a minimal clinically important difference. *npj Primary Care Respiratory Management* 2016; 26: 16041.

[13] Smid, DE, Franssen, FM, Houben-Wilke, S *et al.* Responsiveness and MCID Estimates for CAT, CCQ, and HADS in Patients With COPD Undergoing Pulmonary Rehabilitation: A Prospective Analysis. *Journal of the American Medical Directors Association* 2017; 18(1): 53-58.

[14] Kon, SS, Canavan, JL, Jones, SE *et al.* Minimum clinically important difference for the COPD Assessment Test: a prospective analysis. *Lancet Respiratory Medicine* 2014; 2(3): 195-203.

[15] Dodd, JW, Hogg, L, Nolan, J *et al.* The COPD assessment test (CAT): response to pulmonary rehabilitation. A multicenter, prospective study. *Thorax* 2011; 66(5): 425-429.

[16] Jones, PW. St. George's respiratory questionnaire: MCID. *COPD* 2005; 2(1): 75-79.

[17] Welling, JB, Hartman, JE, Ten Hacken, NH *et al.* The minimal important difference for the St George's Respiratory Questionnaire in patients with severe COPD. *European Respiratory Journal* 2015; 46(6): 1598-1604.

[18] Jones, PW and Bosh, TK. Quality of life changes in COPD patients treated with salmeterol. *American Journal of Respiratory Critical Care Medicine* 1997; 155(4): 1283-1289.

[19] Jaeschke, R, Singer, J and Guyatt, GH. Measurement of health status. Ascertaining the Minimal Clinically Important Difference. *Controlled Clinical Trials* 1989; 10(4): 407-415.

[20] Alma, H, de Jong, C, Tsiligianni, I *et al.* Clinically relevant differences in COPD health status: systematic review and triangulation *European Respiratory Journal* 2018; 52(3): 1800412.

[21] Crosby, RD, Kolotkin, RL and Williams, GR. Defining clinically meaningful change in health-related quality of life. *Journal of Clinical Epidemiology* 2003; 56(5): 395-407.

[22] Revicki, D, Hays, RD, Cella, D *et al.* Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *Journal of Clinical Epidemiology* 2008; 61(2): 102-109.

[23] Beaton, DE, Boers, M and Wells, GA. Many faces of the minimal clinically important difference (MCID): a literature review and directions for future research. *Current Opinion in Rheumatology* 2002; 14(2): 109-114.

[24] Hays, RD and Woolley, JM. The Concept of Clinically Meaningful Difference in Health-Related Quality of Life Research. How Meaningful is it? *Pharmacoeconomics* 2000; 18(5): 419-423.

[25] Raman, S, Ding, K, Chow, E *et al.* Minimal clinically important differences in the EORTC QLQ-BM22 and EORTC QLQ-C15-PAL modules in patients with bone metastases undergoing palliative radiotherapy. *Quality of Life Research* 2016; 25(10): 2535-2541.

[26] Maringwa, J, Quinten, C, King, M *et al.* Minimal clinically meaningful differences for the EORTC QLQ-C30 and EORTC QLQ-BN20 scales in brain cancer patients. *Annals of Oncology* 2011; 22(9): 2107-2112.

[27] Hui, D, Shamieh, O, Paiva, CE *et al.* Minimal Clinically Important Difference in the Physical, Emotional, and Total Symptom Distress Scores of the Edmonton Symptom Assessment System. *Journal of Pain and Symptom Management* 2016; 51(2): 262-269.

[28] Tuomi, L, Johansson, M, Andréll, P and Finizia, C. Interpretation of the Swedish Self Evaluation of Communication Experiences after Laryngeal cancer: Cutoff levels and minimum clinically important differences. *Head Neck* 2016; 38(5): 689-695.

[29] Rouquette, A, Blanchin, M, Sébille, V *et al.* The minimal clinically important difference determined using item response theory models: an attempt to solve the issue of the association with baseline score. *Journal of Clinical Epidemiology* 2014; 67(4): 433-440.

[30] Hui, D, Shamieh, O, Paiva, CE *et al.* Minimal clinically important differences in the Edmonton Symptom Assessment Scale in cancer patients: A prospective, multicentre study. *Cancer* 2015; 121(17): 3027-3035.

[31] Singh, D, D'Urzo, AD, Chuecos, F *et al.* Reduction in clinically important deterioration in chronic obstructive pulmonary disease with aclidinium/formoterol. *Respiratory Research* 2017; 18(1): 106.

[32] Stockley, RA, Edgar, RG, Starkey, S and Turner, AM. Health status decline in α-1 antitrypsin deficiency: a feasible outcome for disease modifying therapies? *Respiratory Research* 2018; 19(1): 137.

[33] Anzueto, AR, Kostikas, K, Mezzi, K *et al.* Indacaterol/glycopyrronium versus salmeterol/fluticasone in the prevention of clinically important deterioration in COPD: results from the FLAME study. *Respiratory Research* 2018; 19(1): 121.

[34] Naya, IP, Tombs, L, Muellerova, H *et al.* Long-term outcomes following first short-term clinically important deterioration in COPD. *Respiratory Research* 2018; 19(1): 222.

[35] Schultz, K, Jelusic, D, Wittmann, M *et al.* Inspiratory muscle training does not improve clinical outcomes in 3-week COPD rehabilitation: results from a randomised controlled trial. *European Respiratory Journal* 2018; 51(1): 1702000.

[36] Kamper, SJ, Maher, C and Mackay, G. Global Rating of Change Scales: A Review of Strengths and Weaknesses and Considerations for Design. *Journal of Manual & Manipulative Therapy* 2009; 17(3): 163-170.

[37] Juniper, EF, Guyatt, GH, Willan, A *et al.* Determining a Minimal Important Change in a Disease-Specific Quality of Life Questionnaire. *Journal of Clinical Epidemiology* 1994; 47(1): 81-87.

[38] Norman, GR, Sloan, JA and Wyrwich, KW. Interpretation of Changes in Health-related Quality of Life. The Remarkable Universality of Half a Standard Deviation. *Medical Care* 2003; 41(5): 582-592.

[39] Terwee, CB, Bot, SD, de Boer, MR *et al.* Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of Clinical Epidemiology* 2007; 60(1): 34-42.

[40] de Vet, HC, Terwee, CB, Ostelo, RW *et al.* Minimal changes in health status questionnaires: distinction between minimally detectable change and minimally important change. *Health and Quality of Life Outcomes* 2006; 4: 54.

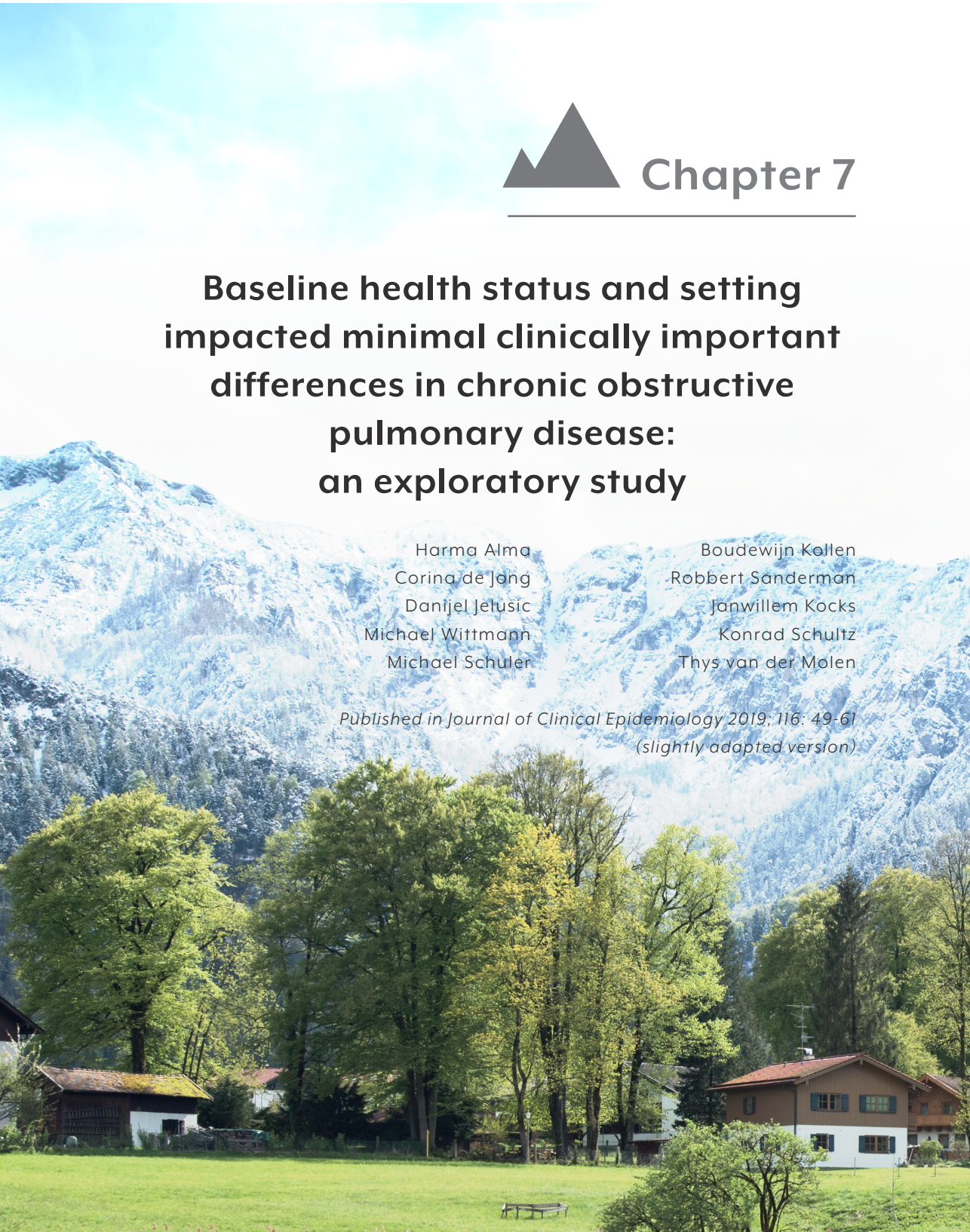
[41] Pothirat, C, Chaiwong, W, Limsukon, A *et al.* Detection of acute deterioration in health status visit among COPD patients by monitoring COPD assessment test score. *International Journal of COPD* 2015; 10: 277-282.

[42] García-Sidro, P, Naval, E, Martinez Rivera, C *et al.* The CAT (COPD Assessment Test) questionnaire as a predictor of the evolution of severe COPD exacerbations. *Respiratory Medicine* 2015; 109(12): 1546-1552.

[43] Janse, AJ, Gemke, RJ, Uiterwaal, CS *et al.* Quality of life: patients and doctors don't always agree: a meta-analysis. *Journal of Clinical Epidemiology* 2004; 57(7): 653-661.

[44] Yoo, JY, Kim, YS, Kim, SS *et al.* Factors affecting the trajectory of health-related quality of life in COPD patients. *International Journal of Tuberculosis and Lung Disease* 2015; 20(6): 738-746.

[45] Sundh, J, Montgomery, S, Hasselgren, M *et al.* Change in health status in COPD: a seven-year follow-up cohort study. *npj Primary Care Respiratory Medicine* 2016; 26: 16073.



Chapter 7

Baseline health status and setting impacted minimal clinically important differences in chronic obstructive pulmonary disease: an exploratory study

Harma Alma
Corina de Jong
Danijel Jelusic
Michael Wittmann
Michael Schuler
Boudewijn Kollen
Robbert Sanderman
Janwillem Kocks
Konrad Schultz
Thys van der Molen

Published in Journal of Clinical Epidemiology 2019; 116: 49-61
(slightly adapted version)

7.1 Abstract

7.1.1 Background

Minimal clinically important differences (MCIDs) are used as fixed numbers in the interpretation of clinical trials. Little is known about its dynamics. This study aims to explore the impact of baseline score, study setting and patient characteristics on health status MCIDs in chronic obstructive pulmonary disease (COPD).

7.1.2 Methods

Baseline and follow-up data on the COPD Assessment Test (CAT), Clinical COPD Questionnaire (CCQ) and St. George's Respiratory Questionnaire (SGRQ) were retrospectively analysed from pulmonary rehabilitation (PR) and routine clinical practice (RCP). Anchor- and distribution-based MCID estimates were calculated and tested between settings, gender, age, global initiative for obstructive lung disease (GOLD) classification, comorbidity level, and baseline health status.

7.1.3 Results

In total, 658 patients were included with 2299 change score measurements. MCID estimates for improvement and deterioration ranged for all subgroups 0.50-6.30 (CAT), 0.10-0.84 (CCQ), and 0.33-12.86 (SGRQ). Larger MCID estimates for improvement and smaller ones for deterioration were noted in patients with worse baseline health status, females, elderly, patients with COPD GOLD grades I/II, and patients with fewer comorbidities. Estimates from PR were generally larger.

7.1.4 Discussion and conclusions

Baseline health status and setting affected MCID estimates of health status questionnaires for patients with COPD. Patterns were observed for gender, age, spirometry classification and comorbidity levels. These outcomes would advocate the need for tailored MCIDs.

7.2 Background

Health status measurement has been defined as *"a standardised way of quantifying the impact of disease on a patient's life, health and well-being"* [1]. It provides important information about the burden of disease - especially in patients with chronic obstructive pulmonary disease (COPD) - because physiologic measures like spirometry alone do not measure the full aspect of this chronic disorder [1-7]. Many factors may impact health status in patients with COPD, including the number and severity of exacerbations [8-10]; the disease severity defined according to the global initiative for obstructive lung disease (GOLD) spirometry classification [5, 9-12]; the patient's gender and age [5, 8-10, 13-15]; and the amount of comorbidities [12]. Moreover, the baseline score may also be predictive of (worsening) health status [11, 14].

Health status is, in addition to spirometry, an obligatory endpoint in evaluating treatment outcomes of clinical trials [16]. An intervention or therapy should result in meaningful clinical changes using outcomes that are relevant for patients themselves [16-18]. A parameter to interpret and define important change in health status is the minimal clinically important difference (MCID). It has been defined as *"the smallest difference in score in the domain of interest which patients perceive as beneficial and which would mandate, in absence of troublesome side effects and excessive costs, a change in patient management"* [19]. Like health status, MCIDs may also be influenced by multiple patient- and disease-related factors. Previous publications speculated that MCIDs for quality of life (QoL) tools may be affected by patient characteristics such as age, gender and education; the number of comorbidities; initial baseline condition of the patient; and pathologic characteristics of the disorder including disease severity [16-17, 20-39]. It remains, however, unclear whether different MCIDs should be used in practice depending on disease severity, patient characteristics and baseline health state.

Various studies have discussed the MCID for health status tools for COPD [40], including the recommended questionnaires COPD Assessment Test (CAT), Clinical COPD Questionnaire (CCQ) and the St. George's Respiratory Questionnaire (SGRQ) [2, 41-42]. The MCIDs were determined as a rather fixed range of 2-3 points on the CAT [40, 43-49], 0.40-0.50 points on the CCQ [40, 43-44, 48-52] and 4-7 points on the SGRQ [40, 43-44, 53-57]. However, little is known about the dynamics of these MCIDs and the impact of different (disease) characteristics of patients on it. The present study aims to explore patient- and disease-related influences upon the MCID estimates for these recommended instruments in patients with COPD from different settings, in this case pulmonary rehabilitation (PR) and routine clinical practice (RCP).

7.3 Methods

7.3.1 Study subjects and design

Data from patients with COPD in two studies were retrospectively analysed: (1) the 3-week full-day inpatient routine inspiratory muscle training within COPD rehabilitation (RIMTCORE) real-life randomised controlled trial during PR in the Klinik Bad Reichenhall (Center for Rehabilitation, Pulmonology and Orthopedics) in Germany [43, 58]; (2) an observational trial (MCID study) aimed primarily at measuring health status changes in Dutch primary and secondary RCP [59].

7.3.2 Data collection

Patient descriptive characteristics and post-bronchodilator spirometry were collected at baseline. CAT (no recall), CCQ (one week recall) and SGRQ (one month recall) were scored at baseline and during follow-up as study outcomes. In the RIMTCORE trial, follow-up questionnaires were scored at discharge (3 weeks) and at 3/6/9/12 months *(Figure 1)*. In RCP, respective health status questionnaires were scored at baseline and after 3/6/12 months. The CAT contains 8 items and a total maximum score of 40 (maximum impairment) [60]. The CCQ consists of 10 items and a total maximum score of 6 (maximum impairment) [61]. The SGRQ contains 50 items and a total score of 100 (maximum impairment) [62]. A 15-point global rating of change (GRC) scale was scored in both studies for each follow-up moment. It required patients to assess their experienced change in health status for COPD compared with baseline. Answers were scored on a scale from -7 (very much worse) to +7 (very much better) with zero equalling no change [63].

7.3.3 Study methods to determine MCIDs

Health status change scores were calculated as the difference between follow-up and baseline scores. Negative change represented improvement, and positive change represented deterioration. MCID estimates were calculated using anchor- and distribution-based methods. The anchor-based approach required patients to be categorised according to their GRC score. Scores of 0 and ±1 on the GRC scale indicated no change; scores of ±2 and ±3 represented minimal improvement/deterioration; scores of ±4 and ±5 were considered moderate improvement/deterioration; and scores of ±6 and ±7 indicated large improvement/deterioration [63]. The mean health status change score of the minimal change group according to the GRC represented the MCID estimate, assuming normality of distribution. In addition, the GRC score for minimally improved and deteriorated was used as cutoff value for dichotomisation in the receiver operating characteristics (ROC) curve analysis [64]. Finally, the distribution-based approach included the half standard deviation (0.5SD) of the health status change score [65].

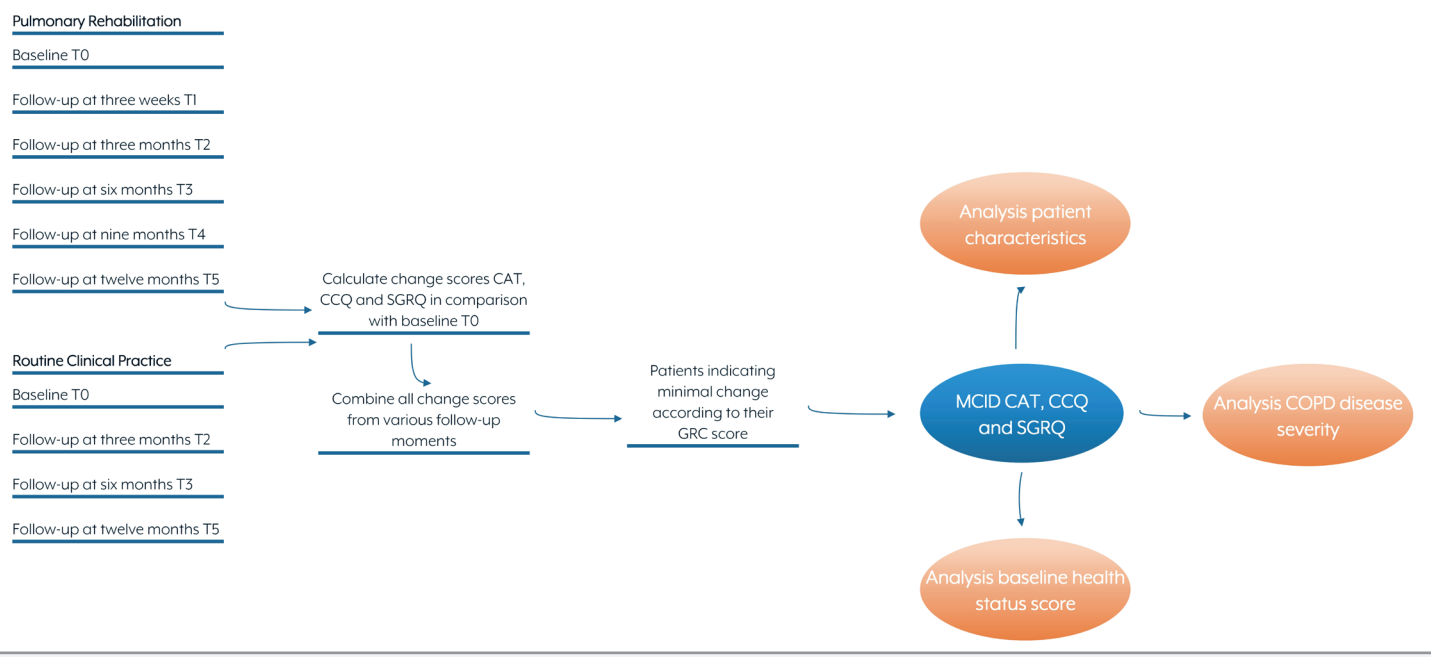
Study methods were performed separately for improvement and deterioration. MCID estimates were evaluated and statistically tested between subgroups for study setting (PR and RCP); males and females; low and high age (median as cutpoint); low (GOLD grade I-II) versus high (GOLD grade III-IV) COPD disease severity; low and high levels of comorbidities on the Charlson Index (median as cutpoint); and better and worse baseline health status (median as cutpoint).

7.3.4 Data analysis and statistics

SPSS version 25.0 (IBM, Chicago, USA) was used for data analysis. Baseline, follow-up and change scores were assessed. Frequencies with percentages, mean with standard deviation (SD), or median with interquartile range (IQR) were determined depending on the variable characteristics. Normality of distribution was assessed for continuous variables using histograms, and skewness and kurtosis with values between -1 and +1 indicative for normality. Baseline scores between both datasets were tested for a difference using independent t-tests, Mann-Whitney U tests or Chi-Square tests depending on type of data and whether assumptions were met. Health status and GRC scores were assessed for floor- and ceiling effects, defined as over 15% of patients scoring in the lowest and highest 10% of the maximum scale range [66]. Baseline and follow-up health status scores were tested for significance of change using paired t-tests or Wilcoxon-signed rank tests. Pair-wise deletion was applied. The false discovery rate (FDR) due to multiple testing was controlled for by Benjamini–Hochberg's stepup procedure to maintain an overall two sided type I error rate of 5% [67]. Adjusted P-values are reported. Intraclass correlation coefficients (ICCs) of the health status change scores were determined to asses dependency of data. The impact of regression to the mean was determined as $100(1-r^{1/2})$ with r reflecting the ICC.

Correlations were assessed between the GRC and health status change scores using Pearson coefficients assuming normality of distribution and the GRC to be a continuous variable. Correlations were required ≥0.30 (preferably ≥0.50) [68]. Patients were categorised per GRC class. Significance of change within each GRC category was tested using paired t-tests or Wilcoxon signed rank tests. ANOVA was performed to statistically test for significant differences between GRC categories and to determine the mean change scores including confidence interval (CI) per class. The mean change scores of the minimally improved and/or deteriorated group of patients represented the MCID estimates. Furthermore, ROC curves used a GRC of ±2 as cutoff value to dichotomise the dataset to differentiate between important and unimportant change [24, 34, 38]. The area under the curve (AUC) was calculated, and the optimal combination of sensitivity and specificity was selected as MCID estimate *(lowest sum of [1-sensitivty] and [1-specificity])* [64].

Figure 1: Study methods of the MCID (subgroup) analysis



Abbreviations: CAT, COPD Assessment Test; CCQ, Clinical COPD Questionnaire; COPD, chronic obstructive pulmonary disease; GRC, global rating of change; MCID, minimal clinically important difference; SGRQ, St. George's Respiratory Questionnaire; T0, baseline; T1, follow-up at 3 weeks; T2, follow-up at 3 months; T3, follow-up at 6 months; T4, follow-up at 9 months; T5, follow-up at 12 months.

Independent t-tests were used to determine significance of the difference between MCID estimates for the various subgroups. These subgroups were determined for nominal values as the respective categories. For the continuous variables, the median was used to dichotomise patients into subgroups [24, 28, 34-35, 69]. To control for the possible impact of the baseline level, relative MCIDs were also calculated as a percentage of change from baseline [21, 29, 36]. Finally, 0.5SDs were determined for all patients as well as the respective subgroups.

Multiple linear regression modeling was applied to quantify the impact of the various factors (baseline health status, gender, age, GOLD classification and study setting) on the MCID estimates, including the analysis of possible interaction terms.

7.4 Results

7.4.1 Baseline characteristics

This retrospective analysis included 451 patients from PR, of whom 309 patients completed follow-up; and 207 patients from RCP, of whom 177 completed follow-up *(Table 1)*.

7.4.2 Global rating of change (GRC)

Correlations between the GRC and health status change scores (n = 2299) were -0.38 (CAT), -0.45 (CCQ) and -0.52 (SGRQ). The scores on the GRC differed significantly between both study settings ($P < 0.001$) *(Figure 2)*. ANOVA tests between GRC groups were all significant ($P < 0.001$).

7.4.3 Study Setting

Mean MCID estimates were for all patients for improvement and deterioration respectively -2.82 vs. 2.66 (CAT), -0.48 vs. 0.43 (CCQ), and -8.30 vs. 5.95 (SGRQ) *(Table 2)*. Estimates were larger during PR compared with RCP for improvement on the CCQ (-0.49 vs -0.40, $P = 0.687$) and SGRQ (-8.71 vs. -3.04, $P = 0.020$); and for deterioration on the CAT (3.44 vs. 1.50, $P = 0.023$), CCQ (0.51 vs. 0.30, $P = 0.090$) and SGRQ (6.11 vs. 5.69, $P = 0.825$) *(Table 2)*. ROC curves demonstrated a similar pattern, although estimates for deterioration were overall smaller than for improvement *(Table 3)*. The 0.5SD estimates were also larger during PR than RCP (CAT 3 vs. 2 points; CCQ 0.5 vs. 0.3 points; and SGRQ 6-7 vs. 4-5 points) *(Table 4)*. ICCs of the health status change scores and percent regression to the mean ranged 0.59-0.68 (17.66-23.12%) for PR and 0.75-0.89 (5.50-13.69%) for RCP.

Table 1: Patient characteristics and health status scores

Variable	Pulmonary Rehabilitation	Routine Clinical Practice	Significance
Number of patients (N)	451	207	
Age (years)[a]	57.87 ± 6.56	66.69 ± 7.91	< 0.001*
Gender (male)[b]	293 (65.0)	121 (58.5)	0.507
FEV_1%pred[a]	50.40 ± 15.11	57.06 ± 21.96	0.012*
GOLD I[b]	-	35 (17.4)	0.239
GOLD II[b]	227 (50.3)	80 (39.8)	
GOLD III[b]	176 (39.0)	61 (30.3)	
GOLD IV[b]	48 (10.6)	25 (12.4)	
Smoking pack years[a]	40 (30-50)	37.5 (22.50-51.25)	0.108
Baseline CAT[a]	20.23 ± 7.33	18.32 ± 7.22	0.003*
Baseline CCQ Total[a]	2.86 ± 1.17	2.12 ± 1.02	< 0.001*
Baseline SGRQ Total[a]	50.69 ± 17.33	42.88 ± 19.16	< 0.001*
Baseline mMRC[a]	2 (2-4)	1 (1-2)	< 0.001*
Δ CAT 12 months[a]	-0.89±7.01	-0.14±4.92	0.224
Δ CCQ Total 12 months[a]	-0.16±1.08	0.02±0.69	0.077
Δ SGRQ Total 12 months[a]	-3.94±15.38	0.87±11.55	< 0.001*

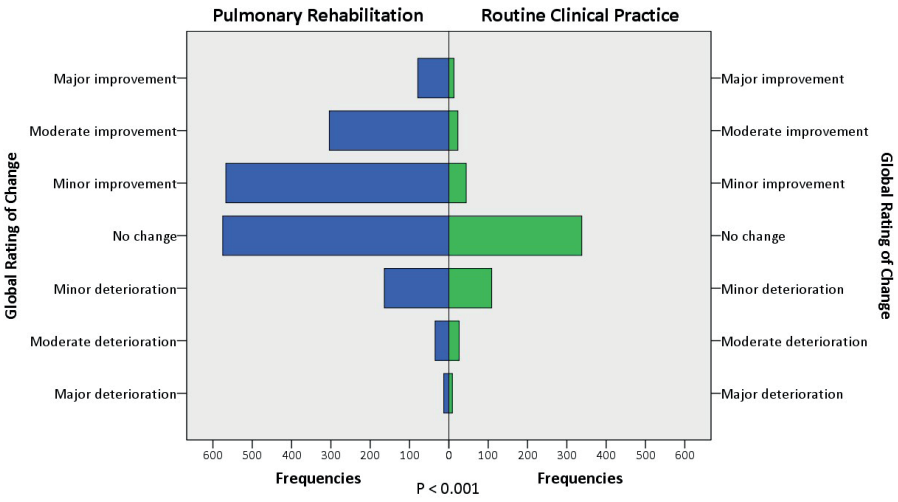
[a] Data expressed as mean ± SD or median (IQR).

[b] Data expressed as frequencies (% of total).

* Significance at level P < 0.05. All listed P-values were corrected using the Benjamini-Hochberg method.

Abbreviations: CAT, COPD Assessment Test; CCQ, Clinical COPD Questionnaire; COPD, chronic obstructive pulmonary disease; FEV_1%pred, forced expiratory volume in one second % predicted; GOLD, global initiative for chronic obstructive lung disease; IQR, interquartile range; mMRC, modified Medical Research Council Dyspnoea scale; N, number of patients; SD, standard deviation; SGRQ, St. George's Respiratory Questionnaire.

Figure 2: Distribution of the GRC scores during PR and RCP



The blue bars (left half) represented the observed frequencies per GRC category during PR. The green bars (right half) represented the observed frequencies per GRC category during RCP. The GRC scores differed significantly between both study settings (P < 0.001).

Abbreviations: GRC, global rating of change; PR, pulmonary rehabilitation; RCP, routine clinical practice.

Table 2: Anchor-based MCID estimates for CAT, CCQ and SGRQ using the mean change method

Patients	N	CAT	CCQ	SGRQ	N	CAT	CCQ	SGRQ
		Improvement	Improvement	Improvement		Deterioration	Deterioration	Deterioration
All Patients	611	-2.82 (-3.30 to -2.35)	-0.48 (-0.55 to -0.41)	-8.30 (-9.32 to -7.27)	272	2.66 (1.99 to 3.32)	0.43 (0.33 to 0.53)	5.95 (4.59 to 7.30)
PR Patients	567	-2.81 (-3.32 to -2.31)	-0.49 (-0.57 to -0.41)	-8.71 (-9.79 to -7.63)	163	3.44 (2.48 to 4.39)	0.51 (0.37 to 0.66)	6.11 (4.16 to 8.06)
RCP Patients	44	-2.91 (-4.27 to -1.55)	-0.40 (-0.58 to -0.21)	-3.04 (-5.52 to -0.57)	109	1.50 (0.70 to 2.30)	0.30 (0.19 to 0.42)	5.69 (3.95 to 7.44)
Significance	611	0.930	0.687	0.020*	272	0.023*	0.090	0.825
Better Baseline Health Status	291 299 317	-0.67 (-1.32 to -0.03)	-0.13 (-0.23 to -0.04)	-4.66 (-6.02 to -3.29)	88 96 95	6.30 (5.07 to 7.52)	0.84 (0.68 to 0.99)	12.86 (10.63 to 15.08)
Worse Baseline Health Status	320 312 290	-4.78 (-5.39 to -4.16)	-0.82 (-0.91 to -0.72)	-12.28 (-13.70 to -10.87)	184 175 173	0.92 (0.26 to 1.58)	0.21 (0.09 to 0.33)	2.15 (0.71 to 3.59)
Significance	611 611 607	< 0.001*	< 0.001*	< 0.001*	272 272 268	< 0.001*	< 0.001*	< 0.001*
Males	370	-2.56 (-3.17 to -1.94)	-0.38 (-0.47 to -0.29)	-8.01 (-9.30 to -6.72)	176	2.74 (1.99 to 3.49)	0.54 (0.42 to 0.67)	6.79 (4.95 to 8.63)
Females	241	-3.23 (-3.97 to -2.48)	-0.63 (-0.75 to -0.51)	-8.74 (-10.43 to -7.06)	96	2.51 (1.20 to 3.82)	0.23 (0.08 to 0.38)	4.43 (2.56 to 6.30)
Significance	611	0.544	0.028*	0.655	272	0.872	0.011*	0.193

Age Low	345	-2.49 (-3.12 to -1.86)	-0.41 (-0.51 to -0.31)	-7.98 (-9.27 to -6.70)	128	3.18 (2.23 to 4.13)	0.50 (0.34 to 0.67)	6.02 (3.83 to 8.22)	
Age High	266	-3.25 (-3.98 to -2.53)	-0.57 (-0.68 to -0.46)	-8.71 (-10.38 to -7.04)	144	2.19 (1.26 to 3.13)	0.37 (0.25 to 0.48)	5.88 (4.19 to 7.56)	
Significance	611	0.413	0.099	0.635	272	0.985	0.370	0.916	
GOLD I-II	310	-3.01 (-3.72 to -2.31)	-0.53 (-0.62 to -0.43)	-9.90 (-11.28 to -8.52)	137	2.35 (1.47 to 3.23)	0.44 (0.29 to 0.59)	4.86 (3.04 to 6.67)	
GOLD III-IV	299	-2.63 (-3.27 to -1.98)	-0.44 (-0.55 to -0.33)	-6.64 (-8.15 to -5.12)	135	2.97 (1.97 to 3.97)	0.42 (0.30 to 0.55)	7.10 (5.07 to 9.13)	
Significance	609	0.661	0.425	0.062	272	0.670	0.976	0.224	
Comorbidities Low	32	-3.16 (-4.89 to -1.43)	-0.41 (-0.63 to -0.19)	-3.37 (-6.54 to -0.20)	64	1.20 (0.14 to 2.27)	0.30 (0.15 to 0.44)	4.42 (2.15 to 6.69)	
Comorbidities High	12	-2.25 (-4.52 to -0.02)	-0.35 (-0.70 to -0.01)	-2.17 (-6.21 to +1.88)	45	1.91 (0.66 to 3.16)	0.32 (0.12 to 0.51)	7.56 (4.81 to 10.31)	
Significance	44	0.778	0.888	0.748	109	0.642	0.941	0.187	

Data presented as mean change scores (95%CI).

Scores marked in lighter green represented smaller MCID estimates; whereas darker green represented larger MCID estimates.

* Significance at level P < 0.05 using independent t-tests. All listed P-values were corrected for multiplicity using the Benjamini-Hochberg method.

Abbreviations: 95%CI, 95% confidence interval; CAT, COPD Assessment Test; CCQ, Clinical COPD Questionnaire; COPD, chronic obstructive pulmonary disease; GOLD, global initiative for chronic obstructive lung disease; N, number of patients; PR, pulmonary rehabilitation; RCP, routine clinical practice; SGRQ, St. George's Respiratory Questionnaire.

Table 3: Anchor-based MCID estimates for CAT, CCQ and SGRQ using ROC Curves

Patients	CAT Improvement	CCQ Improvement	SGRQ Improvement	CAT Deterioration	CCQ Deterioration	SGRQ Deterioration
N	N = 1030	N = 1030	N = 1030	N = 356	N = 356	N = 356
All Patients	-2.50 AUC 0.650	-0.50 AUC 0.693	-7.48 AUC 0.727	+1.50 AUC 0.671	+0.20 AUC 0.679	+4.12 AUC 0.676
PR Patients	-2.50 AUC 0.644	-0.50 AUC 0.687	-7.48 AUC 0.713	+1.50 AUC 0.685	+0.25 AUC 0.661	+4.12 AUC 0.677
RCP Patients	-3.50 AUC 0.650	-0.17 AUC 0.640	-3.26 AUC 0.645	+1.50 AUC 0.651	+0.10 AUC 0.723	+3.57 AUC 0.676
Better Baseline Health Status	-1.50 AUC 0.660	-0.17 AUC 0.673	-6.14 AUC 0.746	+2.50 AUC 0.750	+0.20 AUC 0.751	+7.10 AUC 0.796
Worse Baseline Health Status	-3.50 AUC 0.665	-0.60 AUC 0.743	-9.25 AUC 0.729	+0.50 AUC 0.706	+0.10 AUC 0.699	+0.33 AUC 0.697
Males	-2.50 AUC 0.661	-0.40 AUC 0.686	-6.49 AUC 0.731	+0.50 AUC 0.683	+0.20 AUC 0.689	+4.15 AUC 0.687
Females	-2.50 AUC 0.630	-0.55 AUC 0.697	-7.15 AUC 0.717	+0.50 AUC 0.647	+0.20 AUC 0.660	+1.97 AUC 0.655
Age Low	-1.50 AUC 0.635	-0.50 AUC 0.667	-6.47 AUC 0.710	+0.50 AUC 0.701	+0.25 AUC 0.673	+4.14 AUC 0.692
Age High	-2.50 AUC 0.672	-0.75 AUC 0.724	-7.10 AUC 0.740	+1.50 AUC 0.643	+0.10 AUC 0.686	+3.93 AUC 0.661
GOLD I-II	-2.50 AUC 0.670	-0.60 AUC 0.716	-7.52 AUC 0.765	+0.50 AUC 0.659	+0.10 AUC 0.670	+4.15 AUC 0.669
GOLD III-IV	-1.50 AUC 0.629	-0.20 AUC 0.666	-7.30 AUC 0.683	+0.50 AUC 0.681	+0.10 AUC 0.683	+3.14 AUC 0.677
Comorbidities Low	-3.50 AUC 0.624	-0.50 AUC 0.608	-4.69 AUC 0.628	+0.50 AUC 0.648	+0.20 AUC 0.736	+4.39 AUC 0.641
Comorbidities High	-1.50 AUC 0.710	-0.15 AUC 0.723	-1.78 AUC 0.684	+0.50 AUC 0.643	+0.30 AUC 0.695	+6.03 AUC 0.722

Data presented as MCID estimates from the ROC curves including the AUC.

Scores marked in lighter green represented smaller MCID estimates; whereas darker green represented larger MCID estimates.

Abbreviations: AUC, area under the curve; CAT, COPD Assessment Test; CCQ, Clinical COPD Questionnaire; COPD, chronic obstructive pulmonary disease; GOLD, global initiative for chronic obstructive lung disease; N, number of patients with important change; PR, pulmonary rehabilitation; RCP, routine clinical practice; ROC, receiver operating characteristics; SGRQ, St. George's Respiratory Questionnaire.

Table 4: Distribution-based MCID estimates for CAT, CCQ and SGRQ using the half standard deviation

Patients	CAT Improved	CCQ Improved	SGRQ Improved	CAT Deteriorated	CCQ Deteriorated	SGRQ Deteriorated
All Patients	-2.99	-0.46	-6.43	2.79	0.42	5.65
PR Patients	-3.04	-0.47	-6.54	3.10	0.47	6.28
RCP Patients	-2.24	-0.30	-4.08	2.11	0.30	4.54
Better Baseline Health Status	-2.81	-0.41	-6.17	2.88	0.39	5.46
Worse Baseline Health Status	-2.81	-0.44	-6.12	2.28	0.39	4.80
Males	-3.02	-0.44	-6.30	2.52	0.43	6.12
Females	-2.94	-0.48	-6.63	3.23	0.37	4.62
Age Low	-2.97	-0.47	-6.07	2.71	0.48	6.26
Age High	-3.01	-0.45	-6.89	2.84	0.36	5.07
GOLD I-II	-3.15	-0.44	-6.16	2.61	0.45	5.41
GOLD III-IV	-2.82	-0.49	-6.64	2.95	0.38	5.85
Comorbidities Low	-2.40	-0.31	-4.40	2.13	0.28	4.51
Comorbidities High	-1.79	-0.28	-3.18	2.09	0.33	4.47

Data presented as 0.5SD MCID estimates of the health status change score.

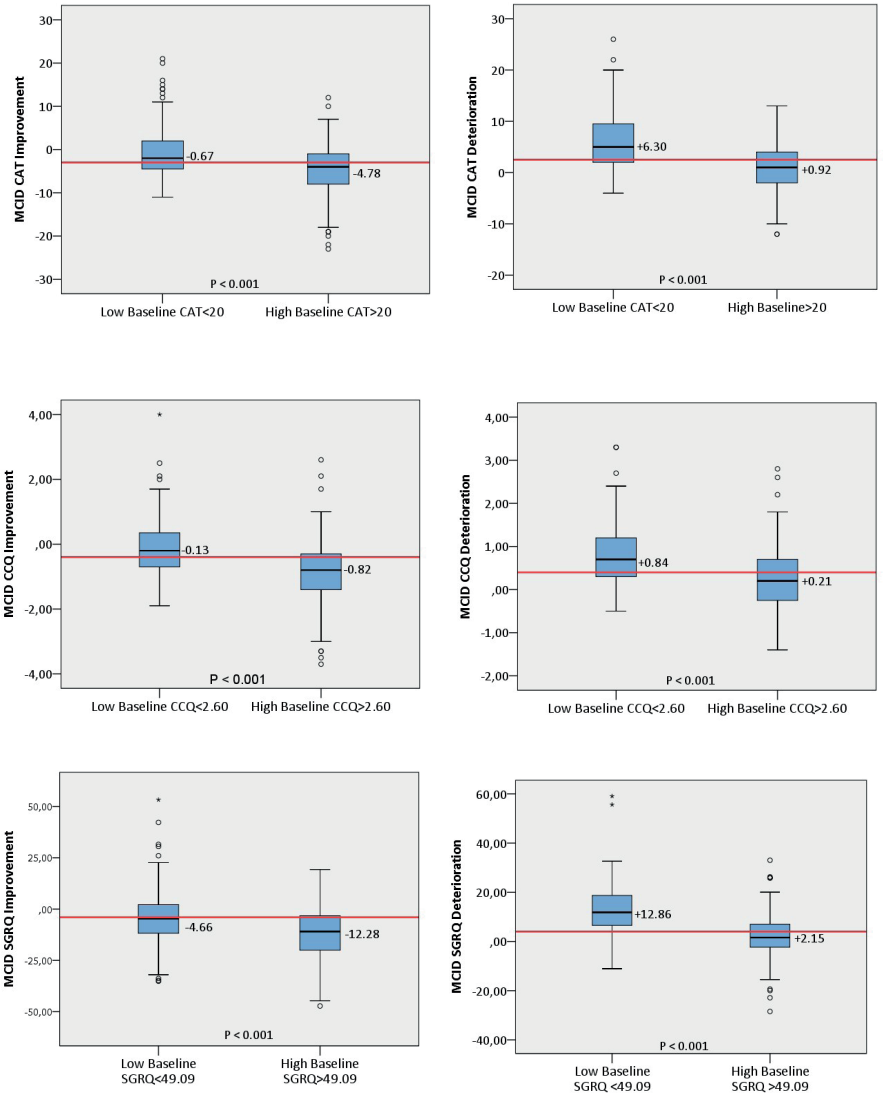
Scores marked in lighter green represented smaller MCID estimates; whereas darker green represented larger MCID estimates.

Abbreviations: 0.5SD, half standard deviation; CAT, COPD Assessment Test; CCQ, Clinical COPD Questionnaire; COPD, chronic obstructive pulmonary disease; GOLD, global initiative for chronic obstructive lung disease; PR, pulmonary rehabilitation; RCP, routine clinical practice; SGRQ, St. George's Respiratory Questionnaire.

7.4.4 Baseline health status

Better and worse baseline levels were grouped according to the median (CAT 20, CCQ 2.60, SGRQ 49.09). MCID estimates for improvement were significantly smaller and for deterioration significantly larger in patients with a lower (*meaning better*) baseline health status during both PR and RCP ($P < 0.001$) (*Table 2, Figure 3, Supplementary material 7.7.1*). Improvement thresholds for lower (*meaning better*) baseline compared with higher (*meaning worse*) baseline were for CAT -0.67 vs. -4.78; for CCQ -0.13 vs. -0.82; for SGRQ -4.66 vs. -12.18. Thresholds for deterioration were for CAT 6.30 vs. 0.92; for CCQ 0.84 vs. 0.21; and for SGRQ 12.86 vs. 2.15. ROC curves confirmed the pattern, although differences were less extreme between baseline severity groups (*Table 3*). The 0.5SD method did not show large differences between both groups (*Table 4, Supplementary material 7.7.2*). Relative MCID estimates compared with baseline were for improvement and deterioration -7.42% and +19.25% (CAT); -11.29% and +22.65% (CCQ), and -14.31% and +18.84% for SGRQ. Their confidence intervals included for improvement -10% and for deterioration +20%. Percentages regression to the mean ranged for low baseline (*meaning better*) patients 31.66-37.55% (PR: 29.36-37.55%; RCP: 38.36-48.52%) and 36.91-39.67% (PR: 38.52-40.84%; RCP: 28.80-34.20%) for high baseline (*meaning worse*) patients.

Figure 3: Box plots of the MCID estimates for both improvement and deterioration defined by low (*meaning better*) and high (*meaning worse*) baseline health status



Box-plots of the MCID estimates for the COPD Assessment Test (CAT), Clinical COPD Questionnaire (CCQ) and St. George's Respiratory Questionnaire (SGRQ) were grouped per health status baseline severity category. The left graphs represented MCID estimates for improvement and the right half represented MCID estimates for deterioration. The red horizontal line represented the currently accepted fixed MCIDs from the literature (CAT 2 points; CCQ 0.40 points; SGRQ 4 points). P-values represented the significant difference between high and low baseline groups.

Abbreviations: CAT, COPD Assessment Test; CCQ, Clinical COPD Questionnaire; COPD, chronic obstructive pulmonary disease; MCID, minimal clinically important difference; SGRQ, St. George's Respiratory Questionnaire.

7.4.5 Other variables

Compared with males, females noted larger MCID estimates for improvement (CAT -3.23 vs. -2.56, P = 0.544; CCQ -0.63 vs. -0.38, P = 0.028; SGRQ -8.74 vs. -8.01, P = 0.655); and smaller ones for deterioration (CAT 2.51 vs. 2.74, P = 0.872; CCQ 0.23 vs. 0.54, P = 0.011; SGRQ 4.43 vs. 6.79, P = 0.193) during both PR and RCP (*Table 2, Supplementary material 7.7.1*). ROC curves partly confirmed this, yet MCID estimates for deterioration were smaller than for improvement (*Table 3*). The 0.5SD estimates were similar for gender, except for deterioration on the CAT and SGRQ (*Table 4, Supplementary material 7.7.2*).

Younger patients had compared with older patients (median 60 years) smaller MCID estimates for improvement (CAT -2.49 vs. -3.25, P = 0.413; CCQ -0.41 vs. -0.57, P = 0.099; SGRQ -7.98 vs. 8.71, P = 0.635); but larger estimates for deterioration (CAT 3.18 vs. 2.19, P = 0.985; CCQ 0.50 vs. 0.37, P = 0.370; SGRQ 5.88 vs. 6.02, P = 0.916) (*Table 2*). This pattern was different in RCP (*Supplementary material 7.7.1*). ROC curves confirmed this pattern, except for deterioration on the CAT (Table 3). The 0.5SD estimates were consistent between both age groups (*Table 4*).

Patients with COPD GOLD grades III-IV scored smaller MCID estimates for improvement (CAT -2.63 vs. -3.01, P = 0.661; CCQ -0.44 vs. -0.53, P = 0.425; and SGRQ -6.64 vs. -9.90, P = 0.062) and larger ones for deterioration (CAT 2.97 vs. 2.35, P = 0.670; SGRQ 7.10 vs. 4.86, P = 0.224) (*Table 2*). The pattern was different for improvement in RCP (*Supplementary material 7.7.1*). ROC curves confirmed the pattern for improvement (*Table 3*). The 0.5SD estimates were consistent between both GOLD groups (*Table 4*).

7.4.6 Linear multiple regression analysis and interaction

Supplementary material 7.7.3 demonstrates the best regression models for the MCID estimates of the CAT, CCQ and SGRQ. Baseline health status and study setting were frequent significant and independent factors in most models. Various interactions were noted. In general, females had worse baseline health status; baseline health status was worse in PR; patients in PR were younger; younger patients had worse baseline health status and GOLD classification; patients with a worse GOLD classification had worse baseline health status. Proportion of the variance of the MCIDs (R^2) explained in these models was between 0.205 and 0.405.

7.5 Discussion

7.5.1 Summary of main findings

The present study demonstrated first of all that MCID estimates for improvement on the CAT, CCQ and SGRQ were significantly three to seven times larger for patients with COPD and a worse baseline health status than for those with better baseline health status; however, they were much smaller for deterioration. Second, MCID estimates proved to be larger during intervention, in this case PR, compared with RCP possibly due to baseline differences. Females, elderly, patients with COPD GOLD grades I and II, and patients with fewer comorbidities had overall larger MCID estimates for improvement and smaller ones for deterioration compared with their counter groups - although not necessarily all significant. Complex interactions between the variables were observed.

7.5.2 Interpretation of outcomes

MCID estimates for CAT, CCQ and SGRQ

Most MCID estimates for the CAT for improvement and deterioration were between ±1.50 and ±3.50, which is in accordance with existing literature [40, 43-49]. MCID estimates for the CCQ were overall in the range of ±0.30 and ±0.60, which also matched previous studies [40, 43-44, 48-52]. SGRQ MCID estimates ranged mostly between 4 and 9 points, which to some extent matched existing evidence [40, 43-44, 53-57]. An estimate of 4 points is generally accepted in interpreting the relevance of clinical trial outcomes. This estimate derived among others from patients treated with Salmeterol in a clinical trial, but also from PR and hospital admitted patients [53-56]. The present study confirms that this 4-point estimate could potentially be valid in RCP, but should be larger for interventions, like PR.

Baseline health status

MCIDs may be dependent on the initial baseline health state of the patient [16-17, 20-21, 23-39, 69]. It did significantly impact our MCID estimates of health status tools for COPD. Higher (*meaning worse*) baseline scores resulted in three to seven times larger MCID estimates for improvement and four to six times smaller estimates for deterioration compared with patients with a better baseline status. This means that these patients required a larger reduction in symptoms and burden of disease before they felt better, and only little deterioration before they felt worse. Interaction between baseline health status and other variables (gender, study setting, age and spirometry classification) was observed, potentially influencing this observed MCID pattern.

Most authors recognise that patients with worse baseline scores require more change before it is to be considered clinically relevant, simply because there is more room for change [16, 20, 24, 30-31, 34-39]. Small improvements are not considered important after intervention or during routine medical care. On the other hand, only small progression of their severe health status would be considered a relevant deterioration. Although perhaps this may be considered a predictable outcome, no former studies in health status for COPD have explored this phenomenon in MCID research. Regression to the mean in our study may explain part of the outcome. Other studies considered the use of relative MCIDs - defined as change in percentage from baseline - to solve the baseline dilemma [20-21, 29, 36, 70]. Relative MCID estimates in the present study were around -10% for improvement and +20% for deterioration. This could possibly be a solution and may perhaps be applied in clinical practice to interpret individual change scores.

COPD disease severity

It has been hypothesised that disease severity - measured in the present study by spirometry classification - could impact the MCID [16-17, 20, 23-39]. Previous - but also present - research demonstrated that worse health status was correlated with worse lung function [9-13], although this correlation was only weak to moderate [4-7]. Patients with worse COPD GOLD grades had also worse baseline health status in our study, and were in general younger. Our study suggested that MCID estimates were larger for improvement and smaller for deterioration in patients with COPD GOLD grades I/II compared with GOLD grades III/IV. This pattern is vice versa the pattern found for the impact of the baseline health status severity on the MCID. Severity of health status is thus not equivalent to COPD disease severity, as expressed in the small to moderate correlation between spirometry and health status. It could be argued that patients with more severe lung function would experience more exacerbations and hospital admissions, which could mean that small changes in the disease state could already be considered important [33]. Age might have interacted in the pattern observed.

Study setting

Setting may impact an instrument's MCID [17, 20, 25, 38], potentially leading to larger MCIDs during intervention [71]. In the current study, MCID estimates for improvement and deterioration in health status for COPD were indeed larger during intervention in PR compared with RCP, although not all results were significant. Patients experienced more change during intervention as a result of treatment, leading to a larger MCID estimate. This perhaps predictable result has not been demonstrated in previous health status research for COPD. A systematic review by Alma *et al.* (2018) [40] on these MCIDs could not observe a similar pattern. In RCP, smaller changes may be noted and regarded relevant. In the present study, patients during PR were significantly younger, and had

worse spirometry and health status at baseline. These factors could have interacted with the different MCID estimates between both settings. The sole impact of setting on the MCID can therefore not be quantified. Furthermore, the sample size during RCP was much smaller, possibly impacting estimates too. Finally, it remains unclear, whether this finding is specific for a rehabilitation intervention or is generally true for any kind of intervention. For this question, further studies would be needed.

Gender, age and comorbidities

Gender, age and comorbidities could impact health status [8-10, 13-15]. First, gender was hypothesised to impact health status and its MCIDs [23]. Men and women evaluate health status differently. Females pay more attention to dyspnoea, emotions and anxiety; and they have more comorbidities [5, 72]. The present study demonstrated that females had (nonsignificant) larger MCID estimates for improvement, and smaller ones for deterioration. Here, the worse baseline health status of females could interact and explain our findings.

Next, age may possibly impact the MCID too [17, 23]. Younger patients experienced significantly worse health status and dyspnoea compared with elderly [5, 73]. However, older age has also been associated with worse health status [5, 14]. In the present study, MCID estimates were larger for improvement and smaller for deterioration in elderly, although not all significant. Older patients had significantly better baseline health status and spirometry. The study in RCP included significantly more elderly than during PR. However, the other interacting patterns found above cannot explain the impact of age on the MCID. Our findings contradict the results by Arima *et al*. [22] that older patients had lower/smaller MCIDs. These authors hypothesised that elderly had lower expectations, being satisfied and thus requiring smaller MCIDs.

Finally, comorbidities could impact health status and its MCIDs [17]. Patients with COPD experience a variety of comorbidities [74]. MCIDs for improvement were larger for patients in RCP with fewer comorbidities, and smaller for deterioration in this group. Comorbidities contribute to the overall disease severity, which would match the pattern found for the COPD disease severity defined by spirometry. Patients with fewer comorbidities were significantly younger, which could imply that age interacted in the pattern too.

7.5.3 Strengths and limitations

The present study is the first study to explore the impact of various factors on MCID estimates for health status tools for COPD. It used a large number of observations obtained from two settings during different follow-up periods applying both anchor- and distribution based approaches. Although the impact of study setting, intervention

and baseline health status score were perhaps predictable, no previous studies have confirmed this in COPD research. No standard approach exists to evaluate the impact of factors on the MCID. The present study dichotomised the impact factors into subgroups. This has been applied by other authors in MCID research outside the field of COPD [24, 28, 34-35, 38-39, 69]. The current authors summed up and subsequently analysed all health status change scores simultaneously to allow for subgroup analysis. The dependency of the change scores was only moderate. Alma *et al*. [44] showed that the recall period was of limited influence on the MCID, supporting the validity of combining all measurements. Finally, the difference between MCID estimates was tested and P-values were corrected for multiple testing.

There are, however, limitations too. Overall, the patterns observed were not all significant after correcting for multiplicity. Owing to the exploratory nature of this study, the current authors chose to report general trends. Next, impact factors on the MCID were analysed individually; however, interactions have been observed between the various factors. This makes simple conclusions difficult to establish; especially as the explained variance R^2 was low. Furthermore, in the anchor-based approach, the GRC was used to differentiate between important and unimportant change. Although correlations fulfilled the pre-set requirements [68], observed correlations were considered weak to moderate. Moreover, it has been argued that GRC estimates could be more related with the follow-up health status score, because of a response shift, therefore not certainly representing change from baseline [75]. Next, the division between PR and RCP patients was unequal, therefore providing more weight to PR measurements. GRC scoring patterns between PR and RCP were also significantly different. Setting impacted the MCID, possibly influencing other subgroup analyses too. Finally, the study on comorbidities, was only valid during RCP, as scores were not readily available for PR.

7.5.4 Implications for clinical practice and future research

At the group level, regression to the mean may play a major role in clinical trials. This means that less weight will be distributed to outlying measurements, balancing out extreme scores. It could be hypothesised that this will minimise the impact of individual patient-related factors and the health status baseline score on the MCID. However, if samples have extreme baseline characteristics or unbalanced divisions, subgroup analyses with clustered MCIDs would be preferred to interpret outcomes in scientific research more precise. The specific trends found for the impact factors on the MCIDs of health status tools for COPD in the present study, might be a start to develop an algorithm in evaluating individual health status changes during clinical practice using tailored MCIDs. More research would be required here to confirm our findings and explore the interactive nature of the variables.

7.5.5 Conclusions

The MCID is currently used as a nonadaptable parameter in the interpretation of health status for patients with COPD in clinical trials. However, our study demonstrated that a complex interaction of study setting, baseline health status, gender, age, spirometry classification and comorbidities potentially impacted the MCID estimates for both improvement and deterioration on three major health status questionnaires. More accurate individual interpretation of outcomes in scientific research and clinical practice would benefit from developing and using possibly clustered or even tailored MCIDs.

7.6 Declarations

7.6.1 Ethics approval and consent to participate

This retrospective study is a secondary analysis of a subsample from the routine inspiratory muscle training within COPD rehabilitation (RIMTCORE) real-life randomised controlled trial (German clinical trial register #DRKS00004609) in the Klinik Bad Reichenhall, Center for Rehabilitation, Pulmonology and Orthopedics in Germany; and a primary analysis of all patients participating in the Dutch observational study on health status in patients with COPD in routine clinical practice (MCID Study; University Medical Center Groningen trial #201500447). All patients in both studies signed informed consent upon participation. The RIMTCORE trial has been approved by the Ethik-Kommission der Bayerischen Landesärztekammer (#12017). The MCID study has been evaluated by the Medical Ethical Committee of the UMCG.

7.6.2 Funding

The main RIMTCORE trial was funded by the Deutsche Rentenversicherung. The Dutch observational study on COPD health status in routine clinical practice (MCID study) as well as the current combined retrospective analysis of both studies received financial support from the Junior Scientific Masterclass as part of the University of Groningen, the Netherlands.

7.6.3 Acknowledgements

We are grateful to the Junior Scientific Masterclass of the University of Groningen, who financially supported the research position of the first author.

7.6.4 Availability of data and material

The data that support the findings of this study are not publicly available. Participating patients in the RIMTCORE trial have only agreed upon availability of their data to the Klinik Bad Reichenhall, their scientific partners in the data analysis and the Committee of

the Bavarian State Chamber of Labor in Munich. Participating patients in the MCID study only agreed upon availability of their data to the UMCG and their scientific partners in the data analysis.

7.6.5 Authors' contributions and consent

Konrad Schultz, Michael Wittmann, Danijel Jelusic and Michael Schuler planned the RIMTCORE study design and were responsible for data collection. Harma Alma, Corina de Jong, Janwillem Kocks, Robbert Sanderman and Thys van der Molen designed the Dutch observational study on COPD health status in routine clinical practice (MCID study) as well as the current retrospective analysis of both studies. Harma Alma, Corina de Jong and Boudewijn Kollen performed the statistical analysis. Harma Alma wrote the first draft, while Corina de Jong, Boudewijn Kollen, Janwillem Kocks, Robbert Sanderman and Thys van der Molen actively participated in the review process. Robbert Sanderman and Thys van der Molen supervised and participated in different steps of the study, as well as in writing. All authors participated in various steps in the study, edited the manuscript and gave their approval for submission.

7.6.6 Competing interests

Harma Alma, Corina de Jong, Danijel Jelusic, Michael Wittmann, Michael Schuler, Boudewijn Kollen and Robbert Sanderman have nothing to disclose. Janwillem Kocks reports personal fees from Novartis; research grants and personal fees from Boehringer Ingelheim; research grants and personal fees from GlaxoSmithKline (GSK); research grants from Stichting Zorgdraad; personal fees from the International Primary Care Respiratory Group (IPCRG); personal fees from Springer Media; and travel arrangements from Chiesi BV, GSK BV, and IPCRG, all outside the submitted work. Konrad Schultz received lecture fees from Boehringer Ingelheim, AstraZeneca, Berlin Chemie, Novartis, Chiesi, Mundipharma, Takeda, GSK and MSD, all outside the submitted work. Thys van der Molen reports personal reimbursements from GSK, TEVA, Astra Zeneca, Boehringer Ingelheim and study grants from Astra Zeneca and GSK. After this study was terminated, he became employee of GSK. None of these prior stated conflicts of interest are linked to the current manuscript. Thys van der Molen developed the CCQ and holds the copyright.

7.7 Supplementary material

7.7.1 Full details anchor-based MCIDs using the mean change method

Table 1: Details anchor-based MCIDs for CAT, CCQ and SGRQ using the mean change method

Patients	N	CAT Improvement	CCQ Improvement	SGRQ Improvement	N	CAT Deterioration	CCQ Deterioration	SGRQ Deterioration
All patients	611	-2.82 ± 5.98 (-3.30 to -2.35)	-0.48 ± 0.92 (-0.55 to -0.41)	-8.30 ± 12.86 (-9.32 to -7.27)	272	2.66 ± 5.57 (1.99 to 3.32)	0.43 ± 0.83 (0.33 to 0.53)	5.95 ± 11.29 (4.59 to 7.30)
PR Patients	567	-2.81 ± 6.08 (-3.32 to -2.31)	-0.49 ± 0.94 (-0.57 to -0.41)	-8.71 ± 13.08 (-9.79 to -7.63)	163	3.44 ± 6.20 (2.48 to 4.39)	0.51 ± 0.94 (0.37 to 0.66)	6.11 ± 12.56 (4.16 to 8.06)
RCP Patients	44	-2.91 ± 4.48 (-4.27 to -1.55)	-0.40 ± 0.59 (-0.58 to -0.21)	-3.04 ± 8.15 (-5.52 to -0.57)	109	1.50 ± 4.22 (0.70 to 2.30)	0.30 ± 0.60 (0.19 to 0.42)	5.69 ± 9.07(3.95 to 7.44)
Significance	611	0.930	0.687	0.020*	272	0.023*	0.090	0.825
Better Baseline Health	291	-0.67 ± 5.61	-0.13 ± 0.82	-4.66 ± 12.34	88	6.30 ± 5.76	0.84 ± 0.77	12.86 ± 10.91
Status All	299	(-1.32 to -0.03)	(-0.23 to -0.04)	(-6.02 to -3.29)	96	(5.07 to 7.52)	(0.68 to 0.99)	(10.63 to 15.08)
	317				95			
Worse Baseline Health	320	-4.78 ± 5.62	-0.82 ± 0.88	-12.28 ± 12.24	184	0.92 ± 4.55	0.21 ± 0.78	2.15 ± 9.59
Status All	312	(-5.39 to -4.16)	(-0.91 to -0.72)	(-13.70 to -10.87)	175	(0.26 to 1.58)	(0.09 to 0.33)	(0.71 to 3.59)
	290				173			
Significance	611	< 0.001*	< 0.001*	< 0.001*	272	< 0.001*	< 0.001*	< 0.001*
	611				272			
	607				268			
Better Baseline Health	269	-0.59 ± 5.68	-0.13 ± 0.86	-5.15 ± 12.49	41	9.17 ± 6.56	1.27 ± 0.91	15.36 ± 12.83
Status PR	310	(-1.09 to -0.09)	(-0.22 to -0.03)	(-6.54 to -3.75)	47	(7.10 to 11.24)	(1.01 to 1.54)	(11.59 to 19.13)
	309				47			
Worse Baseline Health	298	-4.82 ± 5.83	-0.92 ± 0.83	-13.05 ± 12.47	122	1.51 ± 4.73	0.21 ± 0.76	2.33 ± 10.32
Status PR	257	(-5.48 to -4.17)	(-1.03 to -0.82)	(-14.59 to -11.50)	116	(0.66 to 2.36)	(0.07 to 0.35)	(0.43 to 4.24)
	254				115			
Significance	567	< 0.001*	< 0.001*	< 0.001*	163	< 0.001*	< 0.001*	< 0.001*
	567				163			
	563				162			

Better Baseline Health	16	-0.63 ± 4.53	-0.21 ± 0.48	-0.99 ± 8.05	35	4.09 ± 3.22	0.69 ± 0.54	9.96 ± 8.63
Status RCP	22	(-3.04 to +1.79)	(-0.42 to zero)	(-4.39 to +2.41)	35	(2.98 to 5.19)	(0.51 to 0.88)	(7.12 to 12.79)
	24				38			
Worse Baseline Health	28	-4.21 ± 3.96	-0.58 ± 0.65	-5.51 ± 7.74	74	0.27 ± 4.09	0.12 ± 0.55	3.31 ± 8.47
Status RCP	22	(-5.75 to -2.68)	(-0.87 to -0.30)	(-9.15 to -1.88)	72	(-0.68 to 1.22)	(0.01 to 0.24)	(1.26 to 5.36)
	20				68			
Significance	44	0.036*	0.098	<0.188	109	<0.001*	<0.001*	<0.001*
	44				107			
	44				106			
Males All	370	-2.56 ± 6.04 (-3.17 to -1.94)	-0.38 ± 0.88 (-0.47 to -0.29)	-8.01 ± 12.60 (-9.30 to -6.72)	176	2.74 ± 5.04 (1.99 to 3.49)	0.54 ± 0.86 (0.42 to 0.67)	6.79 ± 12.23 (4.95 to 8.63)
Females All	241	-3.23 ± 5.87 (-3.97 to -2.48)	-0.63 ± 0.95 (-0.75 to -0.51)	-8.74 ± 13.26 (-10.43 to -7.06)	96	2.51 ± 6.45 (1.20 to 3.82)	0.23 ± 0.74 (0.08 to 0.38)	4.43 ± 9.23 (2.56 to 6.30)
Significance	611	0.544	0.028*	0.655	272	0.872	0.011*	0.193
Males PR	347	-2.56 ± 6.12 (-3.21 to -1.91)	-0.39 ± 0.90 (-0.48 to -0.29)	-8.38 ± 12.74 (-9.73 to -7.03)	109	3.28 ± 5.37 (2.26 to 4.29)	0.66 ± 0.97 (0.47 to 0.84)	6.50 ± 13.60 (3.89 to 9.10)
Females PR	220	-3.22 ± 6.01 (-4.02 to -2.42)	-0.65 ± 0.98 (-0.78 to -0.52)	-9.23 ± 13.60 (-11.04 to -7.42)	54	3.76 ± 7.66 (1.67 to 5.85)	0.23 ± 0.81 (0.01 to 0.45)	5.36 ± 10.31 (2.58 to 8.15)
Significance	567	0.532	0.028*	0.668	163	0.825	0.016*	0.715
Males RCP	23	-2.52 ± 4.70 (-4.55 to -0.49)	-0.33 ± 0.61 (-0.59 to -0.07)	-2.45 ± 8.77 (-6.24 to +1.35)	67	1.87 ± 4.34 (0.81 to 2.92)	0.35 ± 0.58 (0.21 to 0.50)	7.28 ± 9.66 (4.88 to 9.66)
Females RCP	21	-3.33 ± 4.29 (-5.29 to -1.38)	-0.47 ± 0.59 (-0.73 to -0.20)	-3.69 ± 0.48 (-7.14 to -0.26)	42	0.90 ± 3.99 (-0.34 to +2.15)	0.25 ± 0.64 (0.03 to 0.43)	3.18 ± 7.49 (0.82 to 5.55)
Significance	44	0.816	0.638	0.722	109	0.581	0.496	0.056
Age Low All	345	-2.49 ± 5.93 (-3.12 to -1.86)	-0.41 ± 0.93 (-0.51 to -0.31)	-7.98 ± 12.13 (-9.27 to -6.70)	128	3.18 ± 5.42 (2.23 to 4.13)	0.50 ± 0.94 (0.34 to 0.67)	6.02 ± 12.49 (3.83 to 8.22)
Age High All	266	-3.25 ± 6.02 (-3.98 to -2.53)	-0.57 ± 0.89 (-0.68 to -0.46)	-8.71 ± 13.77 (-10.38 to -7.04)	144	2.19 ± 5.67 (1.26 to 3.13)	0.37 ± 0.71 (0.25 to 0.48)	5.88 ± 10.13 (4.19 to 7.56)
Significance	611	0.413	0.099	0.635	272	0.985	0.370	0.916
Age Low PR	282	-2.48 ± 6.14 (-3.20 to -1.76)	-0.49 ± 0.96 (-0.60 to -0.38)	-8.62 ± 12.68 (-10.11 to -7.13)	98	3.67 ± 5.49 (2.57 to 4.77)	0.57 ± 0.98 (0.37 to 0.77)	6.73 ± 13.29 (4.05 to 9.41)
Age High PR	285	-3.14 ± 6.01 (-3.84 to -2.44)	-0.49 ± 0.91 (-0.60 to -0.38)	-8.80 ± 13.48 (-10.38 to -7.22)	65	3.08 ± 7.17 (1.30 to 4.85)	0.43 ± 0.87 (0.22 to 0.65)	5.19 ± 11.41 (2.36 to 8.02)
Significance	567	0.546	0.989	0.904	163	0.760	0.546	0.735

Age Low RCP	23	-3.65 ± 3.87 (-5.32 to -1.98)	-0.57 ± 0.64 (-0.85 to -0.29)	-5.37 ± 7.79 (-8.74 to -2.00)	47	1.38 ± 3.49 (0.36 to 2.49)	0.33 ± 0.66 (0.14 to 0.52)	4.20 ± 8.17 (1.78 to 6.63)
Age High RCP	21	-2.10 ± 5.03 (-4.00 to -0.19)	-0.20 ± 0.48 (-0.40 to -0.01)	-6.90 ± 13.18 (-15.27 to +1.47)	62	1.58 ± 4.72 (0.38 to 2.78)	0.28 ± 0.56 (0.14 to 0.43)	6.83 ± 9.62 (4.35 to 9.32)
Significance	44	0.547	0.096	0.735	109	0.864	0.890	0.278
GOLD I-II All	310	-3.01 ± 6.30 (-3.72 to -2.31)	-0.53 ± 0.87 (-0.62 to -0.43)	-9.90 ± 12.31 (-11.28 to -8.52)	137	2.35 ± 5.22 (1.47 to 3.23)	0.44 ± 0.90 (0.29 to 0.59)	4.86 ± 10.81 (3.04 to 6.67)
GOLD III-IV All	299	-2.63 ± 5.64 (-3.27 to -1.98)	-0.44 ± 0.96 (-0.55 to -0.33)	-6.64 ± 16.63 (-8.15 to -5.12)	135	2.97 ± 5.90 (1.97 to 3.97)	0.42 ± 0.75 (0.30 to 0.55)	7.10 ± 11.70 (5.07 to 9.13)
Significance	609	0.661	0.425	0.062	272	0.670	0.976	0.224
GOLD I-II PR	288	-3.03 ± 6.45 (-3.78 to -2.29)	-0.53 ± 0.89 (-0.64 to -0.43)	-10.38 ± 12.96 (-11.84 to -8.92)	88	2.93 ± 5.67 (1.73 to 4.13)	0.54 ± 0.98 (0.33 to 0.75)	5.32 ± 11.99 (2.80 to 7.86)
GOLD III-IV PR	279	-2.59 ± 5.67 (-3.26 to -1.92)	-0.44 ± 0.98 (-0.56 to -0.33)	-6.99 ± 13.43 (-8.57 to -5.40)	75	4.03 ± 6.76 (2.47 to 5.58)	0.48 ± 0.85 (0.29 to 0.68)	7.08 ± 13.24 (3.99 to 10.17)
Significance	567	0.669	0.439	0.009*	163	0.524	0.857	0.662
GOLD I-II RCP	22	-2.73 ± 3.84 (-4.43 to -1.02)	-0.40 ± 0.57 (-0.66 to -0.15)	-3.65 ± 6.63 (-6.59 to -0.71)	49	1.31 ± 4.13 (0.12 to 2.49)	0.25 ± 0.61 (0.08 to 0.43)	4.01 ± 8.33 (1.62 to 6.41)
GOLD III-IV RCP	20	-3.15 ± 5.35 (-5.66 to -0.64)	-0.40 ± 0.66 (-0.71 to -0.09)	-1.79 ± 9.78 (-6.37 to +2.79)	60	1.65 ± 4.31 (0.54 to 2.76)	0.35 ± 0.60 (0.19 to 0.50)	7.14 ± 9.50 (4.61 to 9.66)
Significance	42	0.861	1.000	0.658	109	0.858	0.647	0.662
Comorbidities Low RCP	32	-3.16 ± 4.80 (-4.89 to -1.43)	-0.41 ± 0.61 (-0.63 to -0.19)	-3.37 ± 8.79 (-6.54 to -0.20)	64	1.20 ± 4.26 (0.14 to 2.27)	0.30 ± 0.58 (0.15 to 0.44)	4.42 ± 9.01 (2.15 to 6.69)
Comorbidities High RCP	12	-2.25 ± 3.57 (-4.52 to -0.02)	-0.35 ± 0.56 (-0.70 to -0.01)	-2.17 ± 6.36 (-6.21 to +1.88)	45	1.91 ± 4.17 (0.66 to 3.16)	0.32 ± 0.65 (0.12 to 0.51)	7.56 ± 8.94 (4.81 to 10.31)
Significance	44	0.778	0.888	0.748	109	0.642	0.941	0.187

Data presented as mean change score ± SD or mean change score (95% CI).

Scores marked in lighter green represented smaller MCID estimates; whereas darker green represented larger MCID estimates.

* Significance at level P < 0.05 using independent t-tests. All listed P-values were corrected for multiplicity using the Benjamini-Hochberg method.

Abbreviations: 95%CI, 95% confidence interval; CAT, COPD Assessment Test; CCQ, Clinical COPD Questionnaire; COPD, chronic obstructive pulmonary disease; GOLD, global initiative for chronic obstructive lung disease; N, number of patients; PR, pulmonary rehabilitation; RCP, routine clinical practice; SD, standard deviation; SGRQ, St. George's Respiratory Questionnaire.

7.7.2 Full details distribution-based MCIDs

Table 2: Details distribution-based MCID estimates for CAT, CCQ and SGRQ using the half standard deviation

Patients	CAT Improved	CCQ Improved	SGRQ Improved	CAT Deteriorated	CCQ Deteriorated	SGRQ Deteriorated
All patients	-2.99	-0.46	-6.43	2.79	0.42	5.65
PR	-3.04	-0.47	-6.54	3.10	0.47	6.28
RCP	-2.24	-0.30	-4.08	2.11	0.30	4.54
Better Baseline Health Status All	-2.81	-0.41	-6.17	2.88	0.39	5.46
Worse Baseline Health Status All	-2.81	-0.44	-6.12	2.28	0.39	4.80
Better Baseline Health Status PR	-2.84	-0.43	-6.25	3.28	0.46	6.42
Worse Baseline Health Status PR	-2.92	-0.42	-6.24	2.37	0.38	5.16
Better Baseline Health Status RCP	-2.27	-0.44	-4.03	1.61	0.27	4.32
Worse Baseline Health Status RCP	-1.98	-0.33	-3.37	2.05	0.23	4.24
Males All	-3.02	-0.44	-6.30	2.52	0.43	6.12
Females All	-2.94	-0.48	-6.63	3.23	0.37	4.62
Males PR	-3.06	-0.45	-6.37	2.69	0.49	6.80
Females PR	-3.01	-0.49	-6.80	3.83	0.41	5.16
Males RCP	-2.35	-0.31	-4.39	2.17	0.29	4.83
Females RCP	-2.15	-0.30	-3.78	2.00	0.32	3.75
Age Low All	-2.97	-0.47	-6.07	2.71	0.48	6.26
Age High All	-3.01	-0.45	-6.89	2.84	0.36	5.07
Age Low PR	-3.07	-0.48	-6.34	2.75	0.49	6.65
Age High PR	-3.02	-0.46	-6.74	3.59	0.44	5.71
Age Low RCP	-1.94	-0.32	-6.07	2.71	0.33	6.25
Age High RCP	-2.52	-0.24	-6.89	2.84	0.28	5.07
GOLD I-II All	-3.15	-0.44	-6.16	2.61	0.45	5.41
GOLD III-IV All	-2.82	-0.49	-6.64	2.95	0.38	5.85
GOLD I-II PR	-3.23	-0.45	-6.26	2.84	0.51	6.00
GOLD III-IV PR	-2.84	-0.49	-6.72	3.38	0.43	6.62
GOLD I-II RCP	-1.92	-0.29	-3.32	2.17	0.31	4.17
GOLD III-IV RCP	-2.68	-0.33	-4.90	2.16	0.30	4.75
Comorbidities Low RCP	-2.40	-0.31	-4.40	2.13	0.28	4.51
Comorbidities High RCP	-1.79	-0.28	-3.18	2.09	0.33	4.47

Data presented as 0.5SD MCID estimates of the health status change score.

Scores marked in lighter green represented smaller MCID estimates; whereas darker green represented larger MCID estimates.

Abbreviations: 0.5SD, half standard deviation; CAT, COPD Assessment Test; CCQ, Clinical COPD Questionnaire; COPD, chronic obstructive pulmonary disease; GOLD, global initiative for chronic obstructive lung disease; PR, pulmonary rehabilitation; RCP, routine clinical practice; SGRQ, St. George's Respiratory Questionnaire.

7.7.3 Multiple linear regression analysis

Table 3: Multiple linear regression models for the MCIDs of COPD health status instruments

Dependent outcome (MCID)	Best model *excluding* interaction terms	R	R^2	Best model *including* interaction terms	R	R^2
CAT improvement	7.418 – 0.069 (age) + 0.709 (GOLD grade) – 0.406 (baseline CAT)	0.460	0.212	2.967 – 0.007 (interaction age x baseline CAT) + 0.013 (interaction age x GOLD grade)	0.459	0.211
CAT deterioration	16.679 – 3.134 (study setting) – 0.45 (baseline CAT)	0.540	0.291	24.981 – 0.014 (interaction age x baseline CAT) – 20.97 (study setting) + 0.187 (interaction study setting x age) + 0.296 (interaction study setting x baseline)	0.569	0.298
CCQ improvement	0.829 – 0.011 (age) + 0.182 (GOLD grade) – 0.429 (baseline CCQ)	0.499	0.249	0.181 – 0.07 (interaction age x baseline CCQ) + 0.03 (interaction age x GOLD grade) – 0.45 (interaction gender x baseline CCQ)	0.499	0.249
CCQ deterioration	2.698 – 0.637 (study setting) – 0.467 (baseline CCQ)	0.553	0.306	3.718 + 0.248 (interaction setting x baseline CCQ) – 0.816 (baseline CCQ) – 1.324 (study setting)	0.571	0.326
SGRQ improvement	-8.216 + 5.674 (study setting) + 3.831 (GOLD grade) – 0.331 (baseline SGRQ)	0.442	0.195	-12.854 – 0.261 (baseline SGRQ) + 8.233 (GOLD grade) + 0.128 (interaction study setting x baseline SGRQ) – 0.083 (interaction GOLD grade x baseline SGRQ)	0.453	0.205
SGRQ deterioration	31.054 – 5.291 (study setting) + 2.584 (GOLD grade) – 0.448 (baseline SGRQ)	0.614	0.376	51.929 – 0.788 (baseline SGRQ) – 18.613 (study setting) + 0.238 (interaction study setting x baseline SGRQ) + 0.034 (interaction age x GOLD grade)	0.636	0.405

Data presented as the best significant multiple regression model.

Scoring of independent variables: Study setting PR = 1 and RCP = 2; GOLD grade Low = 0 (GOLD I-II) and High = 1 (GOLD III-IV); Baseline health status = continuous score; Gender Male = 0 and Female = 1; Age = continuous score.

Abbreviations: CAT, COPD Assessment Test; CCQ, Clinical COPD Questionnaire; COPD, chronic obstructive pulmonary disease; GOLD, global initiative for obstructive lung disease; MCID, minimal clinically important difference; PR, pulmonary rehabilitation; R, correlation coefficient; R^2, coefficient of determination; RCP, routine clinical practice; SGRQ, St. George's Respiratory Questionnaire.

7.8 References

[1] Jones, PW. Health status measurement in chronic obstructive pulmonary disease. *Thorax* 2001; 56(11): 880-887.

[2] Global initiative for Chronic Obstructive Lung Disease (GOLD). Global Strategy for the Diagnosis, Management and Prevention of COPD (2017 Report). 2017. Available at http://www.goldcopd.org/gold-2017-global-strategy-diagnosis-Management-prevention-copd/. Accessed 9th of July 2018.

[3] Guyatt, GH, Feeny, DH and Patrick, DL. Measuring health-related quality of life. *Annals of Internal Medicine* 1993; 118(8): 622-629.

[4] Westwood, M, Bourbeau, J, Jones, PW *et al*. Relationship between FEV1 change and patient-reported outcomes in randomized trials of inhaled bronchodilators for stable COPD: a systematic review. *Respiratory Research* 2011; 12: 40.

[5] Tsiligianni, I, Kocks, J, Tzanakis, N *et al*. Factors that influence disease-specific quality of life or health status in patients with COPD: a review and meta-analysis of Pearson correlations. *Primary Care Respiratory Journal* 2011; 20(3); 257-268.

[6] Curtis, JR and Patrick, DL. The assessment of health status among patients with COPD. *European Respiratory Journal Supplement* 2003; 41: 36s-45s.

[7] Jones, P, Miravittles, M, van der Molen, T *et al*. Beyond FEV1 in COPD: a review of patient-reported outcomes and their measurement. *International Journal of COPD* 2012; 7: 697-709.

[8] Yoo, JY, Kim, YS, Kim, SS *et al*. Factors affecting the trajectory of health-related quality of life in COPD patients. *International Journal of Tuberculosis and Lung Disease* 2015; 20(6): 738-746.

[9] Rubinsztajn, R, Przybylowski, T, Maskey-Warzechowska, M *et al*. Exacerbations of Chronic Obstructive Pulmonary Disease and Quality of Life of Patients. *Advances in Experimental Medicine, Biology - Neuroscience and Respiration* 2016; 884: 69-74.

[10] Kim, JK, Lee, SH, Lee, BH *et al*. Factors associated with exacerbation in mild-to-moderate COPD patients. *International Journal of COPD* 2016; 11: 1327-1333.

[11] Henoch, I, Strang, S, Löfdahl, C *et al*. Health-related quality of life in a nationwide cohort of patients with COPD related to other characteristics. *European Clinical Respiratory Journal* 2016; 3: 31459.

[12] Wacker, ME, Jörres, RA, Karch, A *et al*. Assessing health-related quality of life in COPD: comparing generic and disease-specific instruments with focus on comorbidities. *BMC Pulmonary Medicine* 2016; 16(1): 70.

[13] Corlateanu, A, Botnaru, V, Covantev, S *et al*. Predicting health-related quality of life in patients with chronic obstructive pulmonary disease: the impact of age. *Respiration* 2016; 92(4): 229-234.

[14] Sundh, J, Montgomery, S, Hasselgren, M *et al*. Change in health status in COPD: a seven-year follow-up cohort study. *npj Primary Care Respiratory Medicine* 2016; 26: 16073.

[15] Nagai, K, Makita, H, Suzuki, M *et al*. Differential changes in quality of life components over 5 years in chronic obstructive pulmonary disease patients. *International Journal of COPD* 2015; 10: 745-757.

[16] Jones, PW, Beeh, KM, Chapman, KR *et al*. Minimal Clinically Important Differences in Pharmacological Trials. *American Journal of Respiratory Critical Care Medicine* 2014; 189(3): 250-255.

[17] Maltenfort, M and Diaz-Ledezma, C. Statistics In Brief: Minimum Clinically Important Difference – Availability of Reliable Estimates. *Clinical Orthopaedics and Related Research* 2017; 475(4): 933-946.

[18] Leopold, SS and Porcher, R. Editorial: The Minimum Clinically Important Difference – The Least We Can Do. *Clinical Orthopedics and Related Research* 2017; 475(4): 929-932.

[19] Jaeschke, R, Singer, J and Guyatt, GH. Measurement of health status. Ascertaining the Minimal Clinically Important Difference. *Controlled Clinical Trials* 1989; 10(4): 407-415.

[20] Angst F, Aeschlimann, A and Angst, J. The minimal clinically important difference raised the significance of outcome effects above the statistical level, with methodological implications for future studies. *Journal of Clinical Epidemiology* 2017; 82: 128-136.

[21] Terwee, CB, Roorda, LD, Dekker, J *et al*. Mind the MIC: large variation among populations and methods. *Journal of Clinical Epidemiology* 2010; 63(5): 524-534.

[22] Arima, H, Carreon, LY, Glassman, SD *et al*. Age variation in the minimum clinically important difference in SRS-22r after surgical treatment for adult spinal deformity – A single institution analysis in Japan. *Journal of Orthopedic Science* 2018; 23(1): 20-25.

[23] Dworkin, RH, Turk, DC, Wyrwich, KW *et al.* Consensus Statement Interpreting the Clinical Importance of Treatment Outcomes in Chronic Pain Clinical Trials: IMMPACT Recommendations. *The Journal of Pain* 2008; 9(2): 105-121.

[24] Wang, YC, Hart, DL, Stratford, PW and Mioduski, JE. Baseline Dependency of Minimal Clinically Important Improvement. *Physical Therapy* 2011; 91(5): 675-688.

[25] Katz, NP, Paillard, FC and Ekman, E. Determining the clinical importance of treatment benefits for interventions for painful orthopedic conditions. *Journal of Orthopedic Surgery and Research* 2015; 10: 24.

[26] Stucki, G, Daltroy, L, Katz, JN *et al.* Interpretation of change scores in ordinal clinical scale and health status measures: the whole may not equal the sum of the parts. *Journal of Clinical Epidemiology* 1996; 49(7): 711-717.

[27] Beaton, DE, Boers, M and Wells, GA. Many faces of the minimal clinically important difference (MCID): a literature review and directions for future research. *Current Opinion in Rheumatology* 2002; 14(2): 109-114.

[28] Baker, DW, Hays, RD and Brook, RH. Understanding Changes in Health Status: Is the Floor Phenomenon Merely the Last Step of the Staircase? *Medical Care* 1997; 35(1): 1-15.

[29] Copay, AG, Subach, BR, Glassman, SD *et al.* Understanding the minimum clinically important difference: a review of concepts and methods. *The Spine Journal* 2007; 7(5): 541-546.

[30] Crosby, RD, Kolotkin, RL and Williams, GR. Defining clinically meaningful change in health-related quality of life. *Journal of Clinical Epidemiology* 2003; 56(5): 395-407.

[31] Hays, RD and Woolley, JM. The Concept of Clinically Meaningful Difference in Health-Related Quality of Life Research. How Meaningful is it? *Pharmacoeconomics* 2000; 18(5): 419-423.

[32] Engel, L, Beaton, DE and Touma, Z. Minimal Clinically Important Difference: a Review of Outcome Measure Score Interpretation. *Rheumatic Disease Clinics North America* 2018; 44(2): 177-188.

[33] Jones, PW, Rennard, S, Tabberer, M *et al.* Interpreting patient-reported outcomes from clinical trials in COPD: a discussion. *International Journal of COPD* 2016; 11: 3069-3078.

[34] Stratford, PW, Binkley, JM, Riddle, DL and Guyatt, GH. Sensitivity to change of the Roland-Morris Back Pain Questionnaire: part 1. *Physical Therapy* 1998; 78(11): 1186-1196.

[35] Schuller, W, Ostelo, RW, Janssen, R and de Vet, HC. The Influence of study population and definition of improvement on the smallest detectable change and the minimal important change of the neck disability index. *Health and Quality of Life Outcomes* 2014; 12: 53.

[36] Farrar, JT, Young, JP Jr, LaMoreaux, L *et al.* Clinical importance of changes in chronic pain intensity measured in an 11-point numerical pain rating scale. *Pain* 2001; 94(2): 149-158.

[37] de Vet, HC, Terwee, CB, Ostelo, RW *et al.* Minimal changes in health status questionnaires: distinction between minimally detectable change and minimally important change. *Health and Quality of Life Outcomes* 2006; 4: 54.

[38] de Vet, HC, Foumani, M, Scholten, MA *et al.* Minimally important change values of a measurement instrument depend more on baseline values than on the type of intervention. *Journal of Clinical Epidemiology* 2015; 68(5): 518-524.

[39] de Vet, HC, Ostelo, RW, Terwee, CB *et al.* Minimally important change determined by a visual method integrating an anchor-based and a distribution-based approach. *Quality of Life Research* 2007; 16(1): 131-142.

[40] Alma, H, de Jong, C, Tsiligianni, I *et al.* Clinically relevant differences in COPD health status: systematic review and triangulation. *European Respiratory Journal* 2018; 52(3): 1800412.

[41] Cave, J, Atkinson, L, Tsiligianni, IG *et al.* Assessment of COPD wellness tools for use in primary care: an IPCRG initiative. *International Journal of COPD* 2012; 7: 447-456.

[42] Kocks, JWH, Seys, SF, van Duin, TS *et al.* Assessing patient-reported outcomes in asthma and COPD patients: which can be recommended in clinical practice? *Current Opinion in Pulmonary Medicine* 2018; 24(1): 18-23.

[43] Alma, H, de Jong, C, Jelusic, D *et al.* Health status instruments for patients with COPD in pulmonary rehabilitation: defining a minimal clinically important difference. *npj Primary Care Respiratory Medicine* 2016; 26: 16041.

[44] Alma, H, de Jong, C, Jelusic, D *et al.* Assessing health status over time: impact of recall period and anchor question on the minimal clinically important difference of copd health status tools. *Health and Quality of Life Outcomes* 2018; 16(1): 130.

[45] Kon, SS, Canavan, JL, Jones, SE *et al.* Minimum clinically important difference for the COPD Assessment Test: a prospective analysis. *Lancet Respiratory Medicine* 2014; 2(3): 195-203.

[46] Dodd, JW, Hogg, L, Nolan, J *et al.* The COPD assessment test (CAT): response to pulmonary rehabilitation. A multicenter, prospective study. *Thorax* 2011; 66(5): 425-429.

[47] Jones PW, Harding, G, Wiklung, I *et al.* Tests of the Responsiveness of the COPD Assessment Test Following Acute Exacerbation and Pulmonary Rehabilitation. *Chest* 2012; 142(1): 134-140.

[48] Smid, DE, Franssen, FM, Houben-Wilke, S *et al.* Responsiveness and MCID Estimates for CAT, CCQ, and HADS in Patients With COPD Undergoing Pulmonary Rehabilitation: A Prospective Analysis. *Journal of the American Medical Directors Association* 2017; 18(1): 53-58.

[49] Tsiligianni, IG, van der Molen, T, Moraitaki, D *et al.* Assessing health status in COPD. A head-to-head comparison between the COPD assessment test (CAT) and the clinical COPD questionnaire (CCQ). *BMC Pulmonary Medicine* 2012, 12: 20.

[50] Canavan, JL, Dilaver, D, Clark, AL *et al.* Clinical COPD Questionnaire in patients with chronic respiratory disease. *Respirology* 2014; 19(7): 1006-1012.

[51] Kocks, JW, Tuinenga, MG, Uil, SM *et al.* Health status measurement in COPD: the minimal clinically important difference of the clinical COPD questionnaire. *Respiratory Research* 2006; 7: 62.

[52] Kon, SS, Dilaver, D, Mittal, M *et al.* The Clinical COPD Questionnaire: response to pulmonary rehabilitation and minimal clinically important difference. *Thorax* 2014; 69(9): 793-798.

[53] Osman, IM, Godden, DJ, Friend, JA *et al.* Quality of life and hospital re-admission in patients with chronic obstructive pulmonary disease. *Thorax* 1997; 52(1): 67-71.

[54] Jones, PW and Bosh, TK. Quality of life changes in COPD patients treated with salmeterol. *American Journal of Respiratory Critical Care Medicine* 1997; 155(4): 1283-1289.

[55] Jones, PW. St. George's respiratory questionnaire: MCID. *COPD* 2005; 2(1): 75-79.

[56] Schünemann, HJ, Griffith, L, Jaeschke, R *et al.* Evaluation of the minimal important difference for the feeling thermometer and the St. George's Respiratory Questionnaire in patients with chronic airflow obstruction. *Journal of Clinical Epidemiology* 2003; 56(12): 1170-1176.

[57] Welling, JB, Hartman, JE, Ten Hacken, NH *et al.* The minimal important difference for the St George's Respiratory Questionnaire in patients with severe COPD. *European Respiratory Journal* 2015; 46(6): 1598-1604.

[58] Schultz, K, Jelusic, D, Wittmann, M *et al.* Inspiratory muscle training does not improve clinical outcomes in 3-week COPD rehabilitation: results from a randomised controlled trial. *European Respiratory Journal* 2018; 51(1): 1702000.

[59] Alma, HJ, de Jong, C, Jelusic, D *et al.* Thresholds for clinically important deterioration versus improvement in COPD health status: results from a randomized controlled trial in pulmonary rehabilitation and an observational study during routine clinical practice. *BMJ Open* 2019; 9(6): e025776.

[60] Jones, PW, Harding, G, Berry, P *et al.* Development and first validation of the COPD Assessment Test. *European Respiratory Journal* 2009; 34(3): 648-654.

[61] Van der Molen, T, Willemse, BWM, Schokker, S *et al.* Development, validity and responsiveness of the Clinical COPD Questionnaire. *Health and Quality of Life Outcomes* 2003, 1: 13.

[62] Jones, PW, Quirck, FH and Baveystock, CM. The St George's Respiratory Questionnaire. *Respiratory Medicine* 1991; 85 (Suppl B): 25-31.

[63] Juniper, EF, Guyatt, GH, Willan, A *et al.* Determining a Minimal Important Change in a Disease-Specific Quality of Life Questionnaire. *Journal of Clinical Epidemiology* 1994; 47(1): 81-87.

[64] de Vet, HC, Terwee, CB, Mokkink, LB and Knol, DL. Measurement in medicine: a practical guide. Cambridge: Cambridge University Press, 2011.

[65] Norman, GR, Sloan, JA and Wyrwich, KW. Interpretation of Changes in Health-related Quality of Life. The Remarkable Universality of Half a Standard Deviation. *Medical Care* 2003; 41(5): 582-592.

[66] Terwee, CB, Bot, SD, de Boer, MR *et al.* Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of Clinical Epidemiology* 2007; 60(1): 34-42.

[67] Benjamini, Y and Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society B* 1995; 57: 289-300.

[68] Revicki, D, Hays, RD, Cella, D *et al.* Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *Journal of Clinical Epidemiology* 2008; 61(2): 102-109.

[69] Rouquette, A, Blanchin, M, Sébille, V *et al*. The minimal clinically important difference determined using item response theory models: an attempt to solve the issue of the association with baseline score. *Journal of Clinical Epidemiology* 2014; 67(4): 433-440.

[70] Zhang, Y, Zhang, S, Thabane, L *et al*. Although not consistently superior, the absolute approach to framing the minimally important difference has advantages over the relative approach. *Journal of Clinical Epidemiology* 2015; 68(8): 888-894.

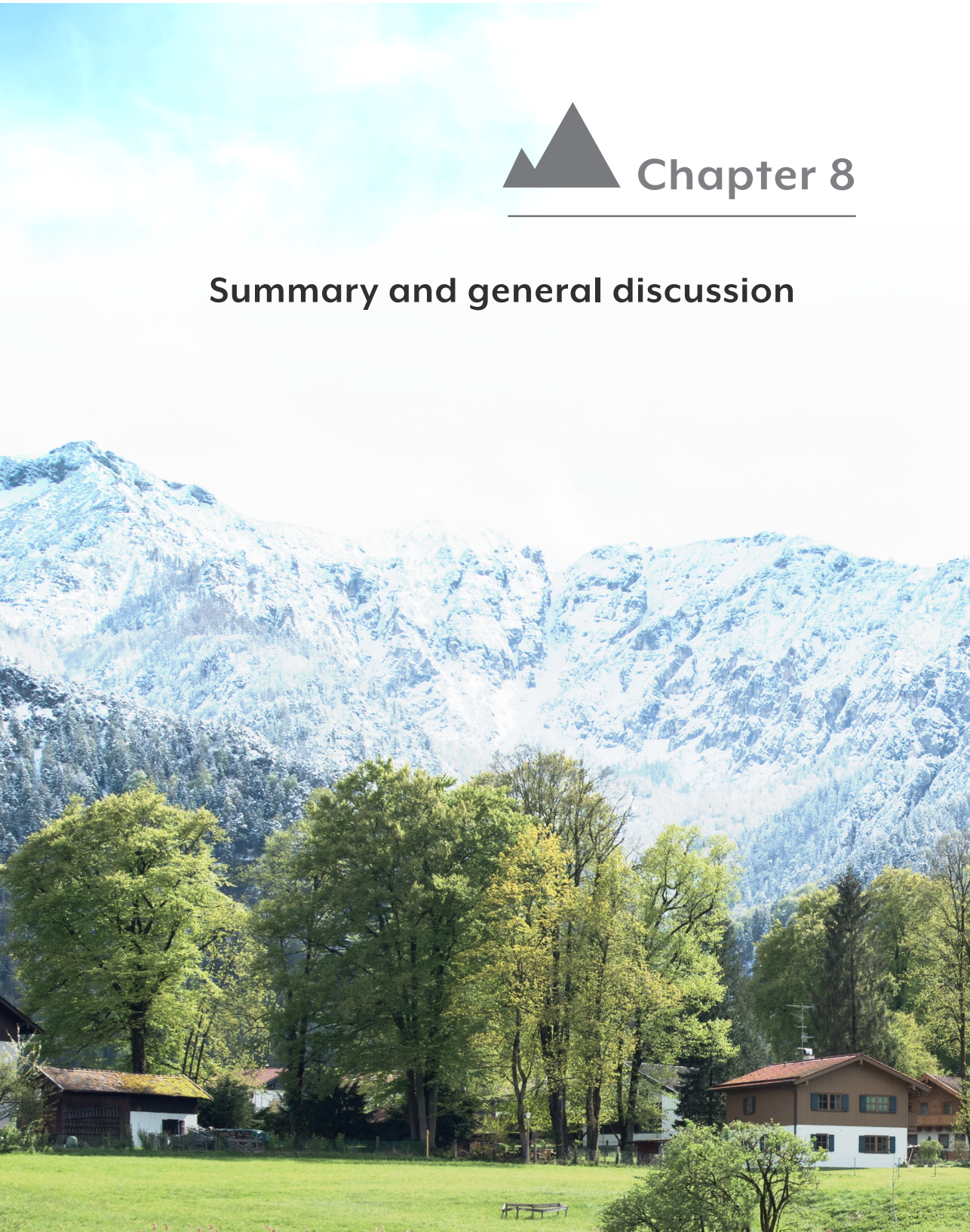
[71] Yuksel, S, Ayhan, S, Nabiyev, V *et al*. Minimum Clinically Important Difference of the Health-Related Quality of Life Scales in Adult Spinal Deformity Calculated by Latent Class Analysis: Is It Appropriate to Use the Same Values for Surgical and Non-Surgical Patients? *The Spine Journal* 2019; 19(1): 71-78.

[72] Nicolini, A, Barbagelata, E, Tagliabue, E *et al*. Gender differences in chronic obstructive pulmonary diseases: a narrative review. *Panminerva Medicine* 2018; 60(4): 192-199.

[73] Martinez, CH, Diaz, AA, Parulekar, AD *et al*. Age-Related Differences in Health-Related Quality of Life in COPD. An Analysis of the COPDGene and SPIROMICS Cohorts 2016. *Chest* 2016; 149(4): 927-935.

[74] Corlateanu, A, Covantev, S, Mathioudakis, AG *et al*. Prevalence and burden of comorbidities in Chronic Obstructive Pulmonary Disease. *Respiratory Investigation* 2016; 54(6): 387-396.

[75] Ousmen, A, Conroy, T, Giullemin, F *et al*. Impact of the occurrence of a response shift on the determination of the minimal important difference in a health-related quality of life score over time. *Health and Quality of Life Outcomes* 2016; 14(1): 167.



Chapter 8

Summary and general discussion

8.1 Background

The primary research objective of this thesis was to gain insight into the dynamics of the *minimal clinically important difference* (MCID) as an outcome parameter of health status instruments used in patients with chronic obstructive pulmonary disease (COPD). The concept of health status measurement was introduced in *Chapter 1*, where it was defined as a standardised way of quantifying and scoring the impact of health and disease on a patient's quality of life (QoL) and well-being [1-7]. In many chronic diseases, including COPD, standard physiological outcome parameters often fail to cover the full impact of the disease on the patient [1-3, 5-19]. Indeed, such tests lack a strong correlation with the patient's experienced QoL, symptoms, functional capabilities and mental status. Therefore, many generic and disease-specific patient-reported outcome (PRO) questionnaires have been introduced during recent decades to quantify the full health and disease states of patients more comprehensively [1-2, 4, 7-13, 20]. These questionnaires have frequently and obligatorily been incorporated as outcome measures in clinical trials (*Box 1*).

Box 1: Interpreting individual results of a clinical trial during pulmonary rehabilitation

In *Chapter 1*, an interview was presented with a 63-year old female patient with grade IV COPD according to the global initiative for obstructive lung disease (GOLD) criteria. At the time, she was participating in a randomised controlled clinical trial (RCT) on the effects of adding inspiratory muscle training (IMT) to an extended 3-week pulmonary rehabilitation (PR) intervention in Germany (*the RIMTCORE trial*). Her main goals were to minimise dyspnoea, prevent the recurrence of exacerbations and slow disease progression. Driving her car again was another major goal.

Immediately after PR, her health status questionnaire scores improved by 4 points on the COPD Assessment Test (CAT), 1.20 point on the Clinical COPD Questionnaire (CCQ) and 10 points on the St. George's Respiratory Questionnaire (SGRQ). However, she experienced an exacerbation during follow-up, which reduced her questionnaire improvements to just 2 points on the CAT, 0.60 point on the CCQ and 7 points on the SGRQ in comparison with her scores prior to PR. Correct interpretation of these improvements is critical to assessing the real effects of PR. Both immediately after PR and during follow-up, it could be argued that this patient experienced clinically relevant improvement, because her change scores exceeded the minimal clinically important differences (MCIDs) calculated in this thesis. Although the MCID is a parameter defined at the group level, as most clinical and physiological parameters are, it can certainly also imply something about individual changes when taking into account a certain level of clinical variation.

In *Chapter 2*, COPD was defined as a chronic disease with multiple, persistent, but not fully reversible, respiratory symptoms caused by obstructive airflow limitation and emphysema [14, 21-22]. Of note, it was highlighted as a leading cause of mortality and morbidity worldwide that had increased over the past decade, which is often present with multiple comorbidities [14, 23-28]. Although there is currently no cure, progression can be halted by combinations of education, self-management, smoking cessation, pharmacotherapy, pulmonary rehabilitation (PR), oxygen therapy and/or surgery [14, 21, 29]. Many tools exist to quantify changes in health status [14, 30-34], but three questionnaires were identified as the most important for patients with COPD: the COPD Assessment Test (CAT, range 0-40), Clinical COPD Questionnaire (CCQ, range 0-6) and the St. George's Respiratory Questionnaire (SGRQ, range 0-100).

In addition to requiring an appropriate questionnaire design for use with COPD, it is important that these tools measure true change (i.e., *a signal*) [6-8, 12, 20, 35]. The observed changes during a trial intervention, such as those reported in *Box 1*, should provide physicians and scientists with valid and clinically important data. This is especially relevant given that many health status PROs have been integrated in scientific research and clinical practice as obligatory measures of the interpretation of change related to an intervention [1-2, 8-12, 36]. The MCID was introduced in *Chapter 1* as the threshold at which this observed change can be considered the minimal that is clinically important and relevant to the patient, thereby justifying the therapy [37-39]. As shown in the case reported in *Box 1*, the parameter allows for the interpretation of change, making it ideal for use as an outcome parameter in clinical trials, and including those related to COPD. However, as identified in *Chapter 1*, there has been a lack of clarity about the dynamics of the MCID in general. In *Chapters 3-7*, the results of measuring the dynamics of the MCID in a COPD population were presented and discussed. A summary of the main findings is provided in the next section, following which there is a discussion that seeks to answer the primary research question, elaborate on the clinical implications and outline opportunities for future research.

8.2 Summary of main findings

The currently accepted MCIDs for the most important health status tools for COPD are 2 points for the CAT, 0.40 point for the CCQ and 4 points for the SGRQ [31-32, 40-42]. To gain insight into the evidence underpinning these MCIDs for the three health status PROs in COPD, a systematic review and triangulation was performed in *Chapter 3* [43]. This identified 21 publications related to the MCIDs of 12 different COPD health status instruments. The results for improvement only (*Table 1, Figures 1-3*) were as follows: -2.54

for the CAT (6 papers, range -3.80 to -1.00), -0.43 for the CCQ (5 papers, range -0.62 to -0.21) and -7.43 for the SGRQ (4 papers, range -10.19 to -2.40). Studies were too few in number and/or too heterogeneous to triangulate the MCIDs for the other nine identified instruments. No patterns were observed for the MCID estimates by setting and/or follow-up period. Of note, evidence for the MCID of the CAT and CCQ was strong and the triangulation seemed valid in comparison with currently accepted MCIDs. However, the MCID currently used for the SGRQ in clinical practice and scientific research did not match that in the reviewed content, which was much higher. Using MCIDs that are too low may have led to an overestimation of treatment effects in clinical trials. Also, MCIDs for deterioration were rarely considered, highlighting the need for more research in this specific domain.

8.2.1 Using multiple approaches in defining a valid MCID

A first important issue regarding the MCID of (COPD) health status tools is that an extensive variety of anchor-, distribution- and opinion-based methods exists to establish an instrument's MCID [44-68]. Some of these are considered better than others, but at present, there is neither a gold standard method nor agreement on the use of each method. In *Chapter 4*, the scales of anchor- and distribution-based methods were investigated with a selection of anchors when determining the MCIDs of the CAT, CCQ and SGRQ for use in COPD [69]. MCIDs for the domain scores of the CCQ and SGRQ were newly determined. The anchor-based methods required change in health status to be compared with another measure of clinical change (e.g., a patient's global rating of change (GRC), a COPD exacerbation as clinical event criterion, or a correlated health status PRO questionnaire) [44, 47, 51-52, 62, 68]. Distribution-based methods required comparison of change with a statistical measure of variability of this change [44, 47, 51-52, 56-57, 62, 64, 66-68]. The resulting MCID estimates differed with the method used in an analysis of 451 patients, who had moderate to very severe COPD (mean age 58 years, 65% male) and participated in a 3-week tailored PR programme with inspiratory muscle training (IMT) as an add-on randomised-controlled intervention [70], giving pooled MCIDs of -3.28 (CAT), -0.52 (CCQ) and -7.91 (SGRQ) (*Table 1, Figures 1-3*). Most MCID estimates for the CAT and CCQ agreed with or were slightly higher than the accepted 2- and 0.40-point thresholds [41-42]. By contrast, all MCID estimates for the SGRQ were higher than the accepted 4-point threshold that has been applied extensively as cutoff value in clinical trials [31-32, 40]. However, its former methodology was analysed and considered unconventional.

Looking at the MCID methods applied in *Chapter 4*, the results from the various anchor-based methods (patient-, criterion- and questionnaire-referencing) were rather comparable. A careful selection of anchors (e.g., *using correlated health status PRO*

questionnaires) was advocated, especially when their MCIDs had not been thoroughly established. The distribution-based method half standard deviation (0.5SD) was best comparable with anchor-based results. The standard error of measurement (SEM) was noted to be inconsistent, while the 1.96SEM was much more conservative for the CAT and SGRQ. Hence, different methods resulted in a range of MCIDs and did not converge on a single estimate. Based upon the pooled MCIDs, however, clinically relevant *improvements* for patients with moderate to very severe COPD during PR would generally be shown at 3 points on the CAT, 0.50 point on the CCQ and 7 points on the SGRQ. These suggested thresholds roughly equate to 7% of the instrument's maximum total score. Domain MCID values were somewhat similar on the CCQ and SGRQ, except for the mental status on the CCQ and symptoms on the SGRQ.

8.2.2 The impact of follow-up length and GRC anchor question on the MCID

A second uncertainty was whether the length of the follow-up period for change measurement affected the MCID estimates due to recall bias or response shift [4, 47, 50, 52, 54, 71-73]. *Chapter 5* summarised a longitudinal exploration of this impact and the design of the GRC anchor question on the MCID of the CAT, CCQ and SGRQ [74]. No significant differences were observed during follow-up at 3 weeks or at 3, 6, 9 and 12 months between the MCID estimates for improvement. Data were evaluated for 451 patients with COPD engaging in a PR intervention of whom 309 completed all follow-up. Using a 15-point GRC anchor question, MCID estimates for improvement ranged from -3.1 to -2.3 for the CAT, -0.6 to -0.4 for the CCQ and -10.3 to -7.6 for the SGRQ (*Table 1, Figures 1-3*). Larger MCIDs were noted for the CAT and CCQ at the 3-week recall period directly after PR, though these were not significantly different. Measuring change over such a short follow-up directly after an event may be associated with a temporary raise in the MCID estimate.

However, MCID results using a 5-point GRC were significantly smaller for the CAT and CCQ in comparison with the 15-point GRC after one year of follow-up. Estimates were -1.4 for the CAT (*significant difference -1.4*), -0.3 for the CCQ (*significant difference -0.2*), and -7.7 for the SGRQ (*nonsignificant difference -1.1*) (*Table 1, Figures 1-3*). Patient classification by both GRCs was only 55% consistent. Hence, introducing fewer choice options on the GRC anchor question resulted in smaller absolute MCIDs. Too few reply options on a GRC could potentially lead to a loss of relevant information, resulting in less discriminative power and lower sensitivity. Overall, the MCIDs for the CAT and CCQ were somewhat comparable or slightly higher than those currently used [41-42], whereas the MCID estimates of the SGRQ were significantly higher compared with those currently used extensively in clinical trials [31-32, 40].

8.2.3 Can we assume similar MCIDs for improvement and deterioration?

A third matter of concern regarding the MCID was whether estimates for clinically relevant improvement compare to those for deterioration [4, 37, 53-54]. As demonstrated in the systematic review in *Chapter 3*, this issue has not previously been explored in the context of COPD [43]. It was therefore addressed in the work presented in *Chapter 6* of this thesis [75]. The primary purpose of an intervention is generally to *improve* health status, but because COPD is a chronic and progressive disease [14, 21-22], preventing *deterioration* should also be considered key outcome of therapy. To do so, one needs to differentiate real worsening of a patient's status (*e.g., signal*) from random variations (*e.g., noise*). Currently, the accepted MCIDs for improvement on the CAT, CCQ and SGRQ are simply also used for deterioration. However, there was no evidence that this approach was appropriate.

When assessing the data at multiple follow-up points for 451 patients during PR and 207 patients during routine clinical practice (RCP), the anchor- and distribution-based MCIDs for improvement and deterioration were similar overall, but the absolute MCIDs differed between the PR and RCP groups (*Table I, Figures I-3*). The results indicated that appropriate cutoff values for both minimal improvement and deterioration could be as follows: CAT ≥3 (*intervention*), CAT ≥2 (*RCP*), CCQ ≥0.40 (*intervention*), CCQ ≥0.30 (*RCP*), SGRQ ≥6-7 (*intervention*) and SGRQ ≥4-5 (*RCP*). Again, the MCID ranges for CAT and CCQ matched well with the currently accepted MCIDs [41-42], but the estimates for the SGRQ during PR were larger than previously reported (although the MCID estimates for the RCP group could be close to the 4-point estimate) [31-32, 40]. Thresholds for moderate and large clinically important improvement and deterioration differed both between each other and between settings; however, the number of patients with moderate or large changes was too small to draw valid conclusions. Thresholds for moderate and large changes need to be explored further, but they may fall in respective ranges of 4-5 and 5-6 points on the CAT, 0.80 and 1.00 point on the CCQ, and 10-15 and 15-20 points on the SGRQ.

8.2.4 The impact of context- and patient-related factors on the MCID

Various factors are known to affect a patient's health status [4, 12, 37, 44, 47, 50, 52, 54-55, 59, 76-79]; however, it is unclear if these also affect the MCID. The impact of patient-related factors, study setting and baseline health status severity on the MCID of the CAT, CCQ and SGRQ were analysed during various follow-up measurements [80]. In total, 658 patients with COPD and no respiratory comorbidities were retrospectively analysed during PR and RCP, providing data for 2299 change scores from baseline. Anchor- and distribution-based methods were applied and the MCID estimates were

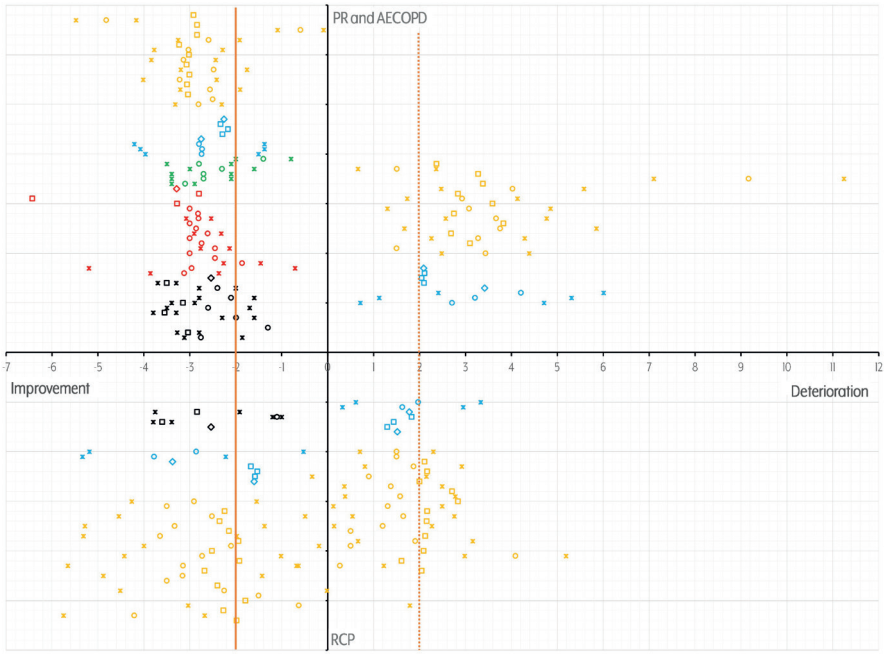
Table 1: Summary of all MCID estimates presented in this thesis

Chapters	CAT		CCQ		SGRQ	
	Improvement	*Deterioration*	*Improvement*	*Deterioration*	*Improvement*	*Deterioration*
Chapter 3 – systematic review [43]	-2.54 (-3.80 to -1.00)	-	-0.43 (-0.62 to -0.21)	-	-7.43 (-10.19 to -2.40)	-
Chapter 4 – various MCID methods [69]	-3.28	-	-0.52	-	-7.91	-
Anchor-based – patient-referencing	-3.12		-0.56		-8.40	
Anchor-based – criterion-referencing	-2.96		-0.62		-9.28	
Anchor-based – questionnaire-referencing	-3.08 to -1.46		-0.61 to -0.28		-9.47 to -6.86	
Distribution-based – SEM	-3.28		-0.29		-5.20	
Distribution-based – 1.96SEM	-6.43		-0.56		-10.19	
Distribution-based – 0.5SD	-2.80		-0.46		-6.06	
Chapter 5 – impact recall period and GRC [74]						
15-point GRC at various follow-up	-3.1 to -2.3	-	-0.6 to -0.4	-	-10.3 to -7.6	-
5 points GRC at follow-up after 12 months	-1.4	-	-0.3	-	-7.7	-
Chapter 6 – improvement and deterioration [75]						
Minimal change – all patients	-3.78 to -1.53	1.30 to 4.21	-0.50 to -0.19	0.19 to 0.66	-9.20 to -2.76	2.75 to 7.53
Minimal change – PR patients	-2.51	2.76	-0.40	0.43	-6.74	5.31
Minimal change – RCP patients	-2.49	1.65	-0.33	0.30	-4.06	4.78
Moderate change – all patients	-4.23	3.89 to 7.06	-0.82	0.62 to 1.23	-16.06	7.46 to 9.30
Chapter 7 – impact factors [80]	-2.82	2.66	-0.48	0.43	-8.30	5.95
Absolute ranges for all subgroups	-6.43 to -0.67	0.50 to 6.30	-0.82 to -0.10	0.19 to 0.84	-12.28 to -2.40	0.33 to 12.86
Most estimates from subgroup analysis	-3.50 to -1.50	1.50 to 3.50	-0.60 to -0.30	0.30 to 0.60	-9.00 to -4.00	4.00 to 9.00

Data from Chapter 3 presented as triangulated MCID (range).
Data from Chapter 4 presented as pooled MCID or MCID estimates/range per method.
Data from Chapters 5, 6 and 7 presented as MCID estimates/range.

Abbreviations: CAT, COPD Assessment Test; CCQ, Clinical COPD Questionnaire; COPD, chronic obstructive pulmonary disease; GRC, global rating of change; MCID, minimal clinically important difference; PR, pulmonary rehabilitation; RCP, routine clinical practice; SD, standard deviation; SEM, standard error of measurement; SGRQ, St. George's Respiratory Questionnaire.

Figure 1: Plot of all MCID estimates obtained for the CAT in this thesis



Overview of all MCID estimates for improvement and deterioration (horizontal axis) from intervention/PR versus RCP (vertical axis). Estimates are presented as circles (anchor-based estimates), squares (distribution-based estimates), diamonds (weighted or triangulated MCIDs), or asterisks (lower and upper limit of the observed range or 95%CI). The vertical orange reference line represented the currently accepted MCID in the literature. The dashed vertical orange reference line mirrors the accepted MCID for improvement into deterioration.

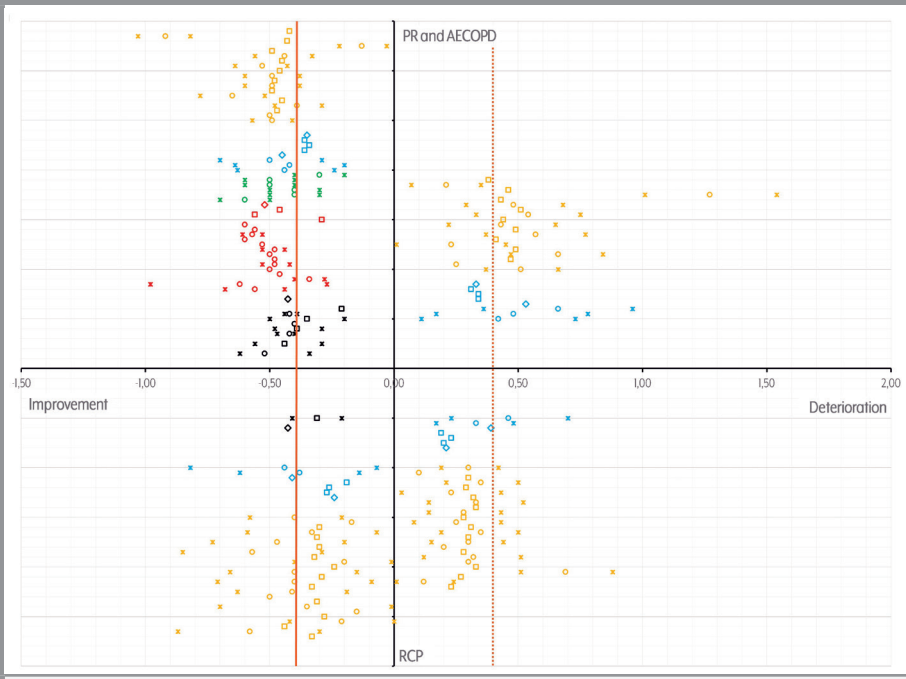
MCID estimates are colour coded per chapter: systematic review and triangulation in Chapter 3 (black); MCID methods in Chapter 4 (red); follow-up period and GRC anchor question in Chapter 5 (green); improvement versus deterioration in Chapter 6 (blue); MCID and dynamic factors in Chapter 7 (orange).

Abbreviations: 95%CI, 95% confidence interval; AECOPD, acute exacerbation chronic obstructive pulmonary disease; CAT, COPD Assessment Test; COPD, chronic obstructive pulmonary disease; MCID, minimal clinically important difference; PR, pulmonary rehabilitation; RCP, routine clinical practice.

evaluated and statistically tested between several subgroups, including study setting (PR vs. RCP), gender (males vs. females), age (median as cutoff), COPD disease severity (GOLD grades I-II vs. GOLD grades III-IV), Charlson Comorbidity Index (median as cutoff) and baseline health status (median as cutoff). Most estimates for improvement and deterioration resulted in a CAT MCID score between ±1.50 and ±3.50, a CCQ MCID score between ±0.30 and ±0.60, and a SGRQ MCID score between ±4 and ±9 points. Trends were also observed during subgroup analyses. Clinically relevant thresholds for change on the respective health status questionnaires differed significantly between the intervention (PR) and RCP groups, with larger absolute MCIDs overall during the intervention. In patients

with worse baseline health statuses, MCID estimates for improvement on the CAT, CCQ and SGRQ were 3-7 times larger; however, MCIDs for deterioration were 4-6 times smaller. Females, older patients (>60 years), patients with GOLD grade I-II COPD by spirometry, and patients with COPD and fewer comorbidities had larger but nonsignificant MCID estimates for improvement, and smaller but nonsignificant estimates for deterioration when compared with their paired subgroup.

Figure 2: Plot of all MCID estimates obtained for the CCQ in this thesis

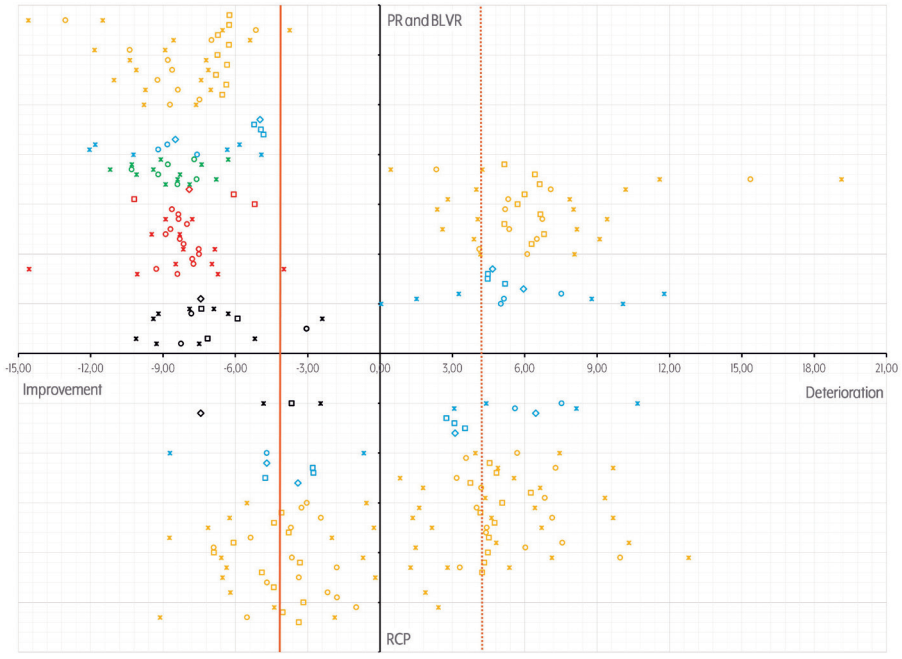


Overview of all MCID estimates for improvement and deterioration (horizontal axis) from intervention/PR versus RCP (vertical axis). Estimates are presented as circles (anchor-based estimates), squares (distribution-based estimates), diamonds (weighted or triangulated MCIDs), or asterisks (lower and upper limit of the observed range or 95%CI). The vertical orange reference line represented the currently accepted MCID in the literature. The dashed vertical orange reference line mirrors the accepted MCID for improvement into deterioration.

MCID estimates are colour coded per chapter: systematic review and triangulation in Chapter 3 (black); MCID methods in Chapter 4 (red); follow-up period and GRC anchor question in Chapter 5 (green); improvement versus deterioration in Chapter 6 (blue); MCID and dynamic factors in Chapter 7 (orange).

Abbreviations: 95%CI, 95% confidence interval; AECOPD, acute exacerbation chronic obstructive pulmonary disease; CCQ, Clinical COPD Questionnaire; COPD, chronic obstructive pulmonary disease; MCID, minimal clinically important difference; PR, pulmonary rehabilitation; RCP, routine clinical practice.

Figure 3: Plot of all MCID estimates obtained for the SGRQ in this thesis



Overview of all MCID estimates for improvement and deterioration (horizontal axis) from intervention/PR versus RCP (vertical axis). Estimates are presented as circles (anchor-based estimates), squares (distribution-based estimates), diamonds (weighted or triangulated MCIDs), or asterisks (lower and upper limit of the observed range or 95%CI). The vertical orange reference line represented the currently accepted MCID in the literature. The dashed vertical orange reference line mirrors the accepted MCID for improvement into deterioration.

MCID estimates are colour coded per chapter: systematic review and triangulation in Chapter 3 (black); MCID methods in Chapter 4 (red); follow-up period and GRC anchor question in Chapter 5 (green); improvement versus deterioration in Chapter 6 (blue); MCID and dynamic factors in Chapter 7 (orange).

Abbreviations: 95%CI, 95% confidence interval; BLVR, bronchoscopic lung volume reduction; MCID, minimal clinically important difference; PR, pulmonary rehabilitation; RCP, routine clinical practice; SGRQ, St. George's Respiratory Questionnaire.

Multiple linear regression models demonstrated that baseline health status severity and setting were frequent significant independent factors in most models of the MCIDs of health status tools used in COPD. Importantly, complex interactions between the variables were observed with possible regression to the mean also noted. The explained variance of the models was low. Although the MCID is currently used as a non-adaptable parameter, there was a complex interaction by study setting, baseline health status, gender, age, spirometry classification and comorbidities, which potentially affected the MCID estimates for both improvement and deterioration on the CAT, CCQ and SGRQ. More accurate individual interpretation of outcomes in scientific research and clinical practice may benefit from developing and using clustered or even tailored MCIDs.

8.3 MCIDs of health status questionnaires in COPD

8.3.1 Important issues in health status interpretation in COPD

To date, the MCID parameter has been used to interpret results in both scientific studies and clinical practice. The SGRQ, in particular, has frequently been used to evaluate health status in patients with COPD. A random example of this can be seen in the study by Gottfried *et al.* [81], who reported the effects of pharmacotherapy with indacaterol 75 µg once daily on dyspnoea and health status. They used the MCID of the SGRQ based on the threshold of 4 points, as has been widely accepted based on prior research [31-32, 40]. In this study, patients using indacaterol improved by 4.9 (group 1) and 5.8 points (group 2) on the SGRQ after 12 weeks. It was concluded that the *within-patient change* exceeded the MCID of 4 points and that the respective odds ratios for achieving the MCIDs of the SGRQ were 1.80 and 1.71. This finding indicated that the *between-patient change* (placebo vs. intervention) based on achieving the MCID favoured the intervention, indicating that indacaterol was an effective therapy for patients with moderate-to-severe COPD.

Many more examples exist in which the MCIDs of health status instruments have been used as important primary outcome parameters to evaluate therapy and interpret its effects in COPD. Given that the approval of new drugs and interventions depends on achieving the MCID, it is pivotal that the parameter is set correctly to prevent inaccurate estimates of treatment effects [77]. As such, it is remarkable that so many different methods are not only used as the basis for the MCID, but that they are also used interchangeably and considered of equal value without clarity on how the outcomes differ [44-68]. These issues result in uncertainties, a lack of clear direction and limited evidence regarding the MCID parameter and its dynamics (*Chapter 1*). Why then, should we as clinicians or researchers be satisfied with the frequent use of the MCID as a key absolute and static evaluation threshold? Indeed, is it now appropriate that we consider the dynamics and complex nature of this parameter? Although any interpretation based on a single MCID threshold appears too simplistic, as currently applied in scientific research and clinical practice, one needs to consider what alternative options realistically exist. Therefore, much of this thesis has focused on investigating the dynamics of the MCID for three recommended PROs used to assess health status in COPD (*I.e., the CAT, CCQ and SGRQ*) [14, 30].

Health status PROs in COPD cover items such as breathlessness, dyspnoea, fatigue, cough, sputum production, physical functioning and exercise tolerance, social functioning, depression and/or anxiety, and exacerbations [13, 18, 82]. Many tools exist to evaluate QoL in COPD (*Chapter 2*). The CAT questionnaire has been recommended by the GOLD committee as the primary health status tool for the evaluation of the symptomatic

burden of COPD in patients [14]. The CCQ, which was developed prior to the CAT, is mostly equivalent, and is perhaps even slightly preferred, because it includes specific domain scores [30, 41, 83-85]. By contrast, the SGRQ is a much more extensive health status PRO that is considered most suitable for use in scientific research [30-32].

The currently accepted MCIDs are 2 points for the CAT, 0.40 point for the CCQ and 4 points for the SGRQ [31-32, 40-42]. The systematic review in *Chapter 3* demonstrated gaps in the evidence for each of these MCIDs [43]. The resulting triangulated values for improvement were -2.54 for the CAT, -0.43 for the CCQ and -7.43 for the SGRQ. Study size, quality and methodology were incorporated in the triangulation to validate its procedures. It was especially worrying that the currently applied MCID of the SGRQ (4 points) [31-32, 40] and of the Chronic Respiratory Questionnaire (CRQ) (MCID 0.5 point on a maximum score of 7 points [38, 86]) were set much lower than the triangulated and reviewed values. Moreover, these lower MCID values were based on evidence that was of poor quality or on studies of questionable value, which in fact were not aimed at determining MCIDs. This may have resulted in overestimation of the treatment effects of many currently approved therapies. The example study on indacaterol by Gottfried *et al.* [81] provides a good example of a pharmacological study with results that exceeded the 4-point MCID threshold for the SGRQ, but that would not reach the triangulated threshold of 7 points. The conclusions of this study should therefore be altered to state that indacaterol might not be an effective therapy for patients with COPD. This conclusion may not be welcomed by all stakeholders.

Given the issues with the SGRQ and CRQ questionnaires, it is alarming that they have been used as anchors to determine the MCIDs of other health status tools and diagnostic tests in COPD too. The CRQ dyspnoea domain, for instance, was used as an anchor in the SGRQ MCID determination process. This resulted in an outlying low estimate of approximately 3 points with insufficient anchor correlations reported [87]. The 4-point MCID of the SGRQ has also been used as an anchor for calculating the MCID of the CAT and CCQ, resulting in these tools having the lowest MCID estimates (*Chapter 4* [69]). In addition to these concerns, the literature review revealed no evidence of MCIDs for deterioration and no research into the impact of study setting or follow-up duration on the MCID estimates (*Chapter 3* [43]). Lastly, it was notable that evidence for the MCID of other COPD health status tools was limited in terms of quantity, quality or both in the systematic review of *Chapter 3* [43]. These MCIDs should not, at this stage, be recommended for use in clinical practice or scientific research for COPD.

8.3.2 MCID outcomes and proposed framework for the dynamic evaluation of changes

The results from the systematic review in *Chapter 3* [43] and the analyses in *Chapters 4–7* [69, 74-75, 80] highlighted that there is a range of MCID values for the CAT, CCQ and SGRQ. In clinical practice, it would be more convenient to use a single cutoff value, although a range does provide more room to consider the methodological-, time-, patient-, direction-, and context-related dynamics of the MCID parameter.

The total ranges for the MCID estimates in this thesis were as follows (*Table I, Figures 1-3)*:

- CAT: -6.43 to -0.67 for *improvement* and 0.50 to 6.30 for *deterioration*;
- CCQ: -0.82 to -0.10 for *improvement* and 0.19 to 0.84 for *deterioration*;
- SGRQ: -12.28 to -2.40 for *improvement* and 0.33 to 12.86 for *deterioration.*

Various outliers were observed in the proposed ranges. Of note, this included the following: the MCID estimates for baseline health status severity subgroups (*Chapter 7* [80]), the 1.96 standard error of measurement (1.96SEM) (*Chapter 4* [69]), the distribution-based estimates of the 95% minimum detectable change (95MDC) and the 0.2 standard deviation (0.2SD) (*Chapter 3* [43]), using the CRQ dyspnoea domain as an anchor for the MCID of the SGRQ (*Chapter 3* [43]), and the estimates obtained from RCP (*Chapters 6 and 7* [75, 80]). Most of the MCID estimates in this thesis [43, 69, 74-75, 80], calculated for both improvement and deterioration, ranged between ±2.00 and ±3.50 for the CAT and between ±0.30 to ±0.50 for the CCQ (*Table I, Figures 1-3)*. These ranges for the SGRQ were -9.00 and -6.00 for improvement and +5.00 to +8.00 for deterioration. Possible cutoff values for clinically relevant moderate and large changes could lie in the range of respectively ±4 to ±5 and ±5 to ±6 points for the CAT, ±0.80 and ±1.00 point for the CCQ, and ±10 to ±15 and ±15 to ±20 points for the SGRQ (*Chapter 6* [75]). These thresholds, however, require more research before they can be recommended with confidence, not least because the number of patients with moderate or large change was small in the study samples. Thresholds for change on the domain scores of the CCQ and SGRQ may also be comparable to those for the total score (C*hapter 4* [69]), but again, more research is required.

Applying the MCID as a range instead of a single cutoff point may have important benefits, potentially offering greater flexibility to the interpretation of change and better incorporating the complexity and dynamics of this parameter [44-47]. One could argue that, based on the findings reported in this thesis thus far, an adaptive extension of the framework as published prior by Man-Son Hing [48] (*Paragraph 1.3.2 Figure 1)* could be suggested. Clinical trial outcomes presented as a point estimate with its 95% confidence interval (95%CI), but also changes observed by physicians in clinical practice, could

be interpreted in relation to the proposed ranges for the CAT, CCQ and SGRQ in this thesis (*Table 2, Figure 4)*. Of course, clinical trial outcomes should at least be statistically significant for any detected change to be considered relevant or important. Then, the outcomes of the scientific trial or in clinical practice should be evaluated in relation to the proposed MCID ranges. If the outcome exceeds the lower threshold of the MCID range, this should be classified as *definite*. However, if the trial outcome does not meet this threshold, but its upper 95%CI does, then a cautious conclusion of *probable or possible* classification may be appropriate.

As example, the study by Gottfried *et al.* on indacaterol once daily [81] could be considered. The trial outcomes (*improvements of 4.9 (group 1) and 5.8 points (group 2) on the SGRQ after 12 weeks*) would classify according to *Table 2* as a definite classification of *"minimal clinicaly important change somewhat likely"*. Based on the value of the upper 95%CI, which unfortunately was not reported, a cautious possible/probable classification of *"minimal clinicaly important change likely"* could be considered for clinical trial outcome on health status.

Important notes must be highlighted when using this proposed classification, with emphasis placed on the importance of researchers and clinicians considering the various dynamic factors of the MCID. The variables of the sample or patient at stake should be incorporated, including the context and baseline variables. Setting can affect the MCID of COPD health status tools, with larger estimates for interventions (in this case PR) compared with stable RCP without additional interventions (*Chapters 6 and 7* [75, 80]). The MCID may also be higher when measuring change directly after intervention/ management compared with measuring change during long-term stable follow-up (*Chapter 5* [74]). Such issues can be incorporated by adjusting the categorisation level. For example, when measuring change during RCP without additional changes in its management, as well as measuring change during long-term follow-up after intervention, it is recommended to step down one level of interpretation in the proposed classification framework (*Table 2, Figure* 4).

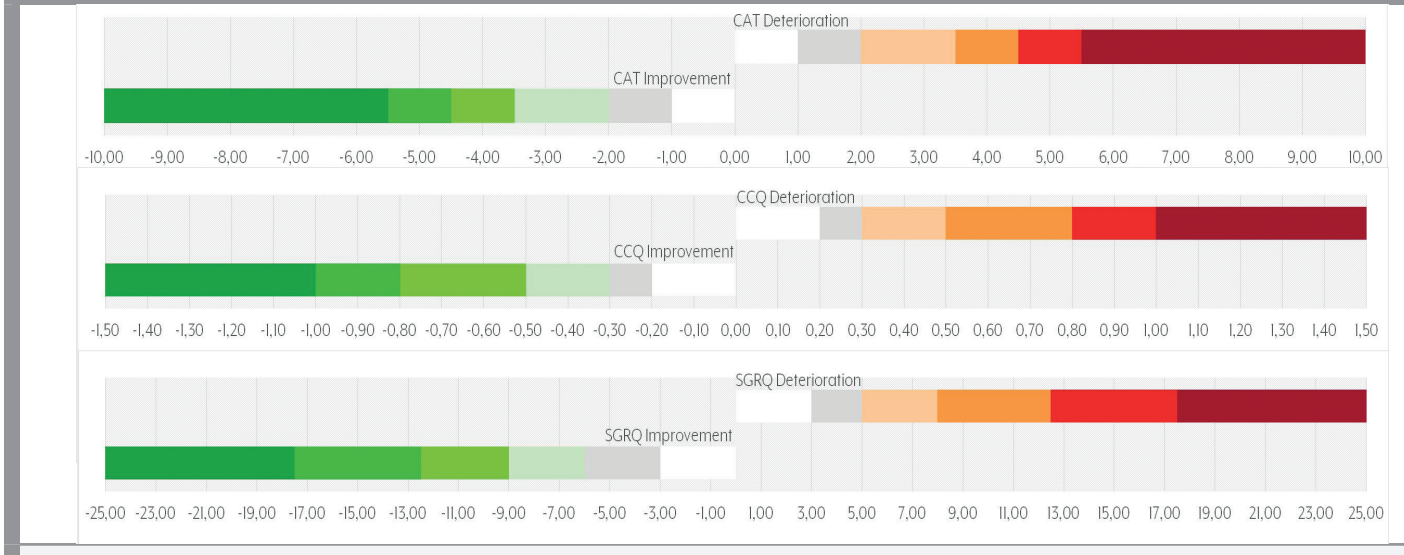
One could also take into account if the investigated cohort includes highly symptomatic patients with worse baseline health status (i.e., CAT ≥20, CCQ ≥2.50 and SGRQ ≥50) and thus require a larger MCID for improvement and a smaller MCID for deterioration. Taking this example further, one may consider shifting categorisation left or right along the horizontal axis of *Figure 4* to alter the strictness of classification according to the baseline health status score. Similar processes could be considered for cohorts with significantly more females and elderly patients (*Chapter 7* [80]), because these subgroups appear to require larger MCIDs for improvement and smaller MCIDs for deterioration too. However,

Table 2: Proposed framework for dynamic MCID evaluation of health status for COPD

	CAT		*CCQ*		*SGRQ*	
	Improvement	*Deterioration*	*Improvement*	*Deterioration*	*Improvement*	*Deterioration*
Minimal clinically important change *unlikely*	> -1.00	< 1.00	> - 0.20	< 0.20	> -3.00	< 3.00
Minimal clinically important change *somewhat likely*	-2.00 to -1.00	1.00 to 2.00	-0.30 to -0.20	0.20 to 0.30	-6.00 to -3.00	3.00 to 5.00
Minimal clinically important change *likely*	-3.50 to -2.00	2.00 to 3.50	-0.50 to -0.30	0.30 to 0.50	-9.00 to -6.00	5.00 to 8.00
Minimal clinically important change *very likely*	< -3.50	> 3.50	< -0.50	> 0.30	< -9.00	> 8.00
Moderate important change *likely*	< -4.50	> 4.50	< -0.80	0.80	< -12.50	> 12.50
Large important change *likely*	< -5.50	> 5.50	<1.00	1.00	< -17.50	> 17.50

Abbreviations: CAT, COPD Assessment Test; CCQ, Clinical COPD Questionnaire; COPD, chronic obstructive pulmonary disease; MCID, minimal clinically important difference.

Figure 4: Proposed framework for dynamic MCID evaluation of health status for COPD



Legend:

- Minimal clinically important change *unlikely*
- Minimal clinically important change *somewhat likely*
- Minimal clinically important improvement/deterioration *likely*
- Minimal clinically important improvement/deterioration *very likely*
- Moderate important improvement/deterioration *likely*
- Large important improvement/deterioration *likely*

The horizontal axes indicate change scores, with improvement (negative values) on the left and deterioration (positive values) on the right. The coloured bars (classified by health status questionnaire) indicate the interpretation category of the observed change for use in scientific trials and clinical practice.

Abbreviations: CAT, COPD Assessment Test; CCQ, Clinical COPD Questionnaire; COPD, chronic obstructive pulmonary disease; MCID, minimal clinically important difference; SGRQ, St. George's Respiratory Questionnaire.

these patient characteristics had less of an impact than either the baseline health status severity or the study setting, and the patterns were less obvious. Interaction may also have influenced these outcomes. The MCID patterns were less clear when samples included patients with worse COPD GOLD grades and/or worse comorbidity levels, which may affect the MCID estimate. More research with larger samples from multiple settings is needed to elucidate these trends, and such research may uncover other influencing factors that should be integrated into the framework for dynamic MCID evaluation (*Table 2, Figure 4*).

8.4 Theoretical considerations in defining an instrument's MCID

Many generic- and disease-specific health status instruments exist [20]. In the future, more tools may emerge or the currently applied tools may be adapted. A central aim of this thesis was to provide clear guidance to help scientists determine an instrument's MCID. Based on the research presented throughout the thesis [43, 69, 74-75, 80], a set of theoretical and practical recommendations has been developed. These recommendations are divided into three quality levels (*Box 2*): bronze, silver and gold. The necessary minimum methodological requirements for determining an MCID are included first, and any additional recommendations to improve the methodological quality and draw final conclusions are included later. Overall, determining an instrument's MCID should combine multiple anchor- and distribution-based methods, preferably obtained from multiple settings with various follow-up periods and predefined usage criteria [44, 47, 49]. The impact of context- and patient-related factors should be incorporated at the next level to investigate the need for clustered MCIDs.

8.4.1 Bronze level – using a variety of MCID methods with sound criteria

The first level of quality (*bronze level*) for determining the MCID is to ensure that appropriate methodology is used (*Box 2 Recommendations 1–5)*. Although there is currently no standard methodology [12, 44, 47, 50], the many different methods that exist can produce a useful range of MCID estimates [4, 12, 39, 44-47, 51-62]. Anchor-based methods use an external reference criterion to provide the threshold for clinically relevant change [4, 12, 39, 44-47, 51-54, 56-61, 79], giving a clear link to medical practice, and often, patients' judgements [63]. Such methods are limited by dependence on correlations between the anchor and the instrument under review and by subjective patient variation [4, 54]. By contrast, distribution-based methods use statistical parameters to assess the significance of a change against its clinical relevance [4, 12, 39, 44-47, 52-53, 56-62, 64]. They can be calculated quickly and can be stable in the face of random variation [4], but they may not convey the clinical importance of an observed change or incorporate patient judgement, and they may result in larger MCIDs in heterogeneous populations.

Differences between MCID estimates obtained by anchor- and distribution-based methods were observed in this thesis too, resulting in the observed range of estimates (*Table 2, Figure 4*). Some good agreement was observed among the various anchor-based methods, though unsurprisingly, with some variation (*Chapter 4* [69]). Using a variety of anchor-based methods with multiple statistical approaches may resolve this issue (*Box 2 Recommendations 1–2).* In the current thesis, criterion-referencing with COPD exacerbations as the clinical criterion event (*so-called between-group change/difference*) was rather equivalent to patient-referencing with a GRC as the anchor (*so-called within-group change*). Given that MCIDs are used for both between and within-patient interpretation of change, this could mean that a similar MCID could be applied in each situation, as follows [36]:

- evaluation of pre- and post-intervention scores *within* the same group;
- evaluation of differences *between* intervention and control groups;
- evaluation of odds ratios between groups in achieving the instrument's MCID after intervention.

It was also shown that the questionnaire-referencing results were comparable to the patient- and criterion-referencing results (*Chapter 4* [69]), despite the major concern that anchor selection was doubtful. Selecting another health status PRO questionnaire as an anchor, especially one for which the MCID has not been thoroughly investigated, could severely impact the MCID of the instrument under review. In *Chapter 4* [69], MCID estimates based on using the original MCID of the SGRQ as the anchor were the lowest of all methods. The MCID for the SGRQ was considered to be set too low. Revised MCIDs based on other anchor-based results for the MCID of the SGRQ better correlated with other methods like patient- and criterion-referencing. The review in *Chapter 3* confirmed the finding that careful anchor selection is needed if one is to guard against faulty MCIDs in new instruments [43]. Clinicians and researchers should therefore avoid using anchors without strong MCIDs set according to the guidelines recommended in this thesis (*Box 2 Recommendation 3).*

In addition to using sound, preferably multiple, anchors and anchor-based approaches (*Box 2 Recommendations 1–3)*, another recommendation is that correlations between the anchors and instruments should be both reported and of an appropriate degree (*Box 2 Recommendation 4*). In the systematic review (*Chapter 3*), it was observed that correlations between selected anchors and studied health status tools not only were infrequently reported, but also were infrequently sufficient [43]. As a criterion for eligibility, correlation coefficients (r) should preferably be ≥0.50, with a minimum standard of ≥0.30 [47]. However, not all correlations between selected anchors and health status instruments in this thesis met the 0.50 threshold, with correlations being lowest directly after an intervention, especially for GRC anchors. Correlations between anchors and change

scores approached or even exceeded the 0.50 threshold during follow-up (*Chapters 5–7* [74-75, 80]). Also, the correlations between anchors and health status questionnaires were stronger for the follow-up health status score than for its change value or baseline score (*Chapter 6* [40]). This may represent a possible response shift [88] that should be taken into consideration (*Box 2 Recommendation 4)*.

In *Chapters 3–7*, multiple statistical methods were used within the anchor-based approaches when calculating the MCID [43, 69, 74-75, 80]. These included the mean change method, the use of receiver operating characteristics (ROC) curves and regression analysis to define thresholds for clinically relevant change. In general, the results were comparable and stable between methods, though it must be noted that using ROC curves resulted in somewhat larger estimates for improvement and smaller estimates for deterioration (*Chapters 4 and 7* [69, 80]). Furthermore, despite requiring the area under the ROC curve (AUC) to be ≥0.70, this was infrequently reported in the reviewed literature (*Chapter 3*) [43]. Multiple statistical approaches are recommended when using anchor-based methods (*Box 2 Recommendation 2)*.

In comparison with the anchor-based approaches, the distribution-based methods appeared more stable over time, and between subgroups and study settings in this thesis. However, these estimates were generally smaller than for the anchor-based outcomes during follow-up (*Chapters 5–7* [74-75, 80]). Distribution-based outcomes could therefore better define the lowest MCID without conveying clinical importance (*Box 2 Recommendation 5*). Anchor-based methods provide real patient and/or clinical perspectives and can provide supportive information [12, 47, 49, 59]. The distribution-based 0.5SD method was most comparable to the anchor-based methods. However, it was less sensitive for evaluating and measuring the dynamics of the MCID over various follow-up periods and between subgroups (*Chapters 5 and 7* [74, 80]). Most 0.5SD results were also lower than those given by anchor-based approaches (*Chapter 6* [75]), especially during RCP (but also during PR). The other distribution-based applications, like the SEM and 1.96 SEM, were less consistent in the explored literature (*Chapter 3*) and throughout the thesis (*Chapter 4*) [43, 69]. Although the 1.96 SEM was much more conservative than the other methods, this has not been reported in previous studies [43]. Due to these inconsistencies, use of the SEM cannot be recommended, contrary to earlier reports [66-67].

Box 2: Theoretical and methodological recommendations for determining an instrument's MCID

Bronze level

Recommendation 1: Combine multiple anchor- and distribution-based methods, including various types of patient-referencing, criterion-referencing and questionnaire-referencing with more than one anchor.

Recommendation 2: Apply various statistical techniques in the anchor-based methods, including the mean change method, ROC curves (including AUCs) and regression analysis.

Recommendation 3: Select anchors for questionnaire-referencing that have well-established MCIDs according to the current recommendation levels here.

Recommendation 4: Assess and report correlations between the selected anchors, including GRCs and health status questionnaires, and requiring a minimum level of $r \geq 0.30$, but preferably $r \geq 0.50$. Assess correlations between the anchor and the instrument's change/baseline/follow-up scores to assess for a potential response shift.

Recommendation 5: Assess the minimum level for clinically relevant change and confirm the anchor-based MCID estimates by calculating the distribution-based 0.5SD in different study settings, populations, measurement periods and subgroups.

Silver level

Recommendation 6: Determine an instrument's MCID over multiple recall periods directly after an intervention and during long-term follow-up. Compare significant differences in estimates for dependency-adjusted confidence intervals based upon the intraclass correlation coefficient (ICC).

Recommendation 7: For patient-referencing in the anchor-based method, select at least two GRC scales during follow-up, with each having a different number of reply categories in both numerical and verbal versions.

Recommendation 8: Determine MCIDs for improvement and deterioration separately with a variety of methods and in different time periods. Do not assume that each MCID type is similar.

Recommendation 9: Expand on *Recommendation 8* to include multiple settings, interventions and routine medical care. If context- or population-specific differences exist, consider using different MCIDs in clinical practice and scientific research.

Gold level

Recommendation 10: Assess the impact of baseline health status on an instrument's MCID by (dichotomised) subgroup analysis and multiple linear regression modelling. If significantly large differences exist, different MCIDs should be used in clinical practice and scientific research when samples/patients have extreme baseline characteristics or when individual assessment of change is required.

Recommendation 11: Assess the impact of appropriate patient- and disease-related factors on an instrument's MCID. If significantly large differences exist, different MCIDs should be used in clinical practice and scientific research when samples/patients have extreme baseline characteristics or when individual assessment of change is required.

Recommendation 12: Assess the *interaction* between patient-, disease- and context-related factors that affect the MCID by independent t-tests and multiple linear regression modeling.

Recommendation 13: Collect all MCID evidence in a systematic review and use triangulation procedures based on the study size, methodological quality, and MCID estimates. Consider whether a single estimate, multiple estimates or a range for dynamic MCID evaluation is required to interpret therapy effects.

The three levels of quality for MCID determination are shown.

Abbreviations: 0.5SD, half standard deviation; AUC, area under the curve; GRC, global rating of change; ICC, intraclass correlation coefficient; MCID, minimal clinically important difference; R, correlation coefficient; ROC, receiver operating characteristics; SEM, standard error of measurement.

8.4.2 Silver level – integrating follow-up duration, direction of change and context

Whereas the bronze level requires that MCIDs are determined using sound anchor- and distribution-based approaches with valid criteria, the second level (*silver level*) requires that researchers look deeper into specific dynamic factors that could affect the MCID of an instrument. Change during various follow-up periods should be investigated (*Box 2 Recommendation 6*), because MCIDs may be influenced by the length of the follow-up (i.e., the *recall period*) [4, 47, 50, 52, 54, 71-73]. The longer the recall period, the more difficult it may be for a patient to assess the change in health status (*recall bias*) [71-73, 88-92], and this indicates that MCIDs can change over time. The systematic review in *Chapter 3* found no specific time-related pattern [43], and testing the hypothesis in *Chapter 5* confirmed that the overall impact of the recall period and of recall bias on the MCID was limited during follow-up [74]. It is possible that the fixed recall moment before PR, as well as the expected symptom stability after PR, limited the expected impact of the recall period.

It was interesting that the MCIDs determined directly after the PR intervention were larger than those measured during follow-up (*Chapter 5)* [74]. This means that MCIDs determined directly after an intervention could be larger than those determined during longer follow-up. At the same time, correlations between the anchor and the reviewed instrument were lowest directly after the intervention, which might help explain the findings. The fact that the calculated MCID was larger after therapy is not surprising: most change would be expected directly after the intervention due to a combination of the therapy itself, placebo effect and/or (high) patient expectations. MCIDs are typically used when evaluating change directly after an intervention, though interpreting change during follow-up may need to be an additional goal when evaluating long-term trial effects. Even though there was only a limited impact of recall bias on the MCID in this thesis, other authors have shown that the recall period affects patients' assessments of change [54, 71-72, 91-97]. It is therefore recommended to check if the recall period affects the MCID estimate (*Box 2 Recommendation 6*).

GRCs are often used as anchors when measuring change during follow-up (so-called *patient-referencing)*. Their use is recommended at the bronze level in *Box 2* (*Recommendation 1*). However, at the silver level, it is important to investigate the impact of the design of the GRC anchor question on the MCID (*Box 2 Recommendation 7*). Fewer choice options with the GRC resulted in less discriminative power and lower MCID estimates in this thesis (*Chapter 5* [74]). There was only limited agreement (55%) between anchor GRCs with 5 and 15 choice options, so patients scored the GRCs differently. This may have been due to the number of choice categories, but may also have been due to the design (i.e., verbal vs. numerical scales). In the systematic review (*Chapter 3* [43]), it was noted that GRCs with fewer response categories were frequently used, possibly explaining the smaller MCID estimates seen for the SGRQ and (possibly) for the CAT.

A third recommendation at the silver level is that the direction of change should be taken into consideration (i.e., improvement and deterioration; *Box 2 Recommendation 8*). When doing so, one should certainly consider all prior recommendations. MCIDs have usually been developed for use with interventions designed to improve health status, which is an important goal of therapy. However, in a chronic progressive disease like COPD, preventing true deterioration is an important outcome too [11, 98-100]. Measuring and preventing deterioration would perhaps be of most interest in routine medical care and for patients with worse baseline disease severity. The systematic review of *Chapter 3* clearly demonstrated that MCIDs for deterioration were non-existent in the existing literature for COPD [43], a situation that may be reflected in other medical sciences.

MCIDs for improvement and deterioration are not necessarily similar [4, 37, 47, 53-54]. While some studies found differences between these MCIDs, others did not [101-107]. Also, it was hypothesised elsewhere that MCIDs for improvement would be smaller than those for deterioration [4, 107], but this was not confirmed in the current thesis. Overall, no major structural discrepancies between MCIDs for improvement and deterioration were observed during PR and RCP [75, 80], although some differences were noted with the CAT during RCP and the SGRQ during PR (*Chapter 7* [80]); in those cases, MCIDs for deterioration were smaller than those for improvement. By contrast, the MCID of the SGRQ during RCP was larger for deterioration than for improvement, possibly due to a difference in the number of patients in the respective change groups. Relative MCIDs were also smaller for improvement (approximately 10% change from baseline) than for deterioration (approximately 20% change from baseline) (*Chapter 7* [80]). Generally smaller absolute estimates were also observed in *Chapter 7* for deterioration than for improvement when comparing subgroups by baseline health status severity, gender, age, GOLD grade and comorbidity [80]. Thus, one should not assume that MCIDs for improvement and deterioration are alike (*Box 2 Recommendation 8*). Thresholds for moderate and large deterioration were also notably different from those for improvement [75]; however, interpretation of these data was complicated by the rather small number of patients in these categories.

A final silver recommendation is that study setting should be considered for the MCIDs for both improvement and deterioration (*Box 2 Recommendation 9*). As has already been stated, differences were shown to exist between improvement and deterioration measures during PR and RCP, while measures directly after PR were associated with higher MCID thresholds (*Chapter 5 [74]*). Importantly, the study setting could also influence these results, especially considering that MCIDs appear to be context- or setting specific [12, 37, 47, 52, 59, 79, 78, 108]. In *Chapters 6 and 7,* it was shown that larger MCIDs for both improvement and deterioration (except for the CAT during RCP) tended to be observed during PR compared with RCP [75, 80]. This was clear with both anchor- and distribution-based methods. Possible baseline differences in health status disease severity, age and spirometry could have influenced the observed trend. No structural pattern for study setting was observed in the systematic review (*Chapter 3* [43]), but it is still considered important that multiple settings and populations be explored to check for significant differences in the MCID threshold, and that this should include both intervention and routine medical care groups.

8.4.3 Gold level – integrating patient-, disease- and context-related factors

At the highest level of MCID determination (*gold level*), efforts should be made to incorporate patient-, disease- and context-related factors into the MCID framework (*Box 2 Recommendations 10 and 11*). Many factors demonstrated to affect health status, potentially also influencing the MCID for health status tools. Baseline- and patient-related factors have previously been hypothesised to influence MCID estimates [12, 36-37, 44, 50, 52, 54, 68, 76-77, 105, 107, 109-119], but there was a lack of evidence prior to this thesis. In general, large differences were found to exist in MCID estimates for patients with a worse baseline health status score than in those with a better baseline score (*Chapter 7* [80]). MCIDs for improvement were 3- to 7-times *larger* for patients with worse baseline health statuses, while MCIDs for deterioration were 4- to 6-times *smaller* for patients with worse baseline health statuses. This reflected the fact that patients with worse baseline health status simply had greater room for improvement and less room for deterioration. One could also conclude that patients with an already severe health status only require a small deterioration to feel worse and much larger improvement to feel better. The study setting could have affected this pattern too, because the PR intervention trial included more patients with severe baseline health status scores.

Overall, females also had *larger* MCIDs for improvement and *smaller* MCIDs for deterioration. Although females generally interpret health status differently to males [120], it should be noted that females had a worse baseline health status in the study in this thesis. Nonsignificant trends were furthermore noted for age, disease severity and comorbidity levels. MCIDs for improvement were *larger* in groups comprising older patients, those with better spirometry results and those with fewer comorbidities. By contrast, their MCIDs were generally *smaller* for deterioration. Younger patients had worse overall baseline health statuses and were more represented in the PR intervention, which contradicts the observed results for age.

Chapter 7 highlighted the difficulty of simple interpretation when defining the impact of single factors on an instrument's MCID [80]. Given that a complex pattern of interaction interrupted the process, it is recommended that factors be considered in an integrated manner for MCID determination (*Box 2 Recommendation 12*). However, the best regression models in this thesis could not explain more than 20%–40% of the observed variation in the MCID. Furthermore, phenomena such as regression to the mean could have occurred at the group level, especially in the analysis of baseline health status severity: in large samples, any baseline and patient-related differences tend to disappear at the group level. However, in samples with extreme characteristics or when more individual or clustered analyses are needed, integrating factors could be worthwhile for considering the effects of therapy. Therefore, a dynamic framework for (individual) MCID interpretation was proposed in *Paragraph 8.3*.

At the pinnacle of the gold level for determining an instrument's MCID, it is recommended that all evidence should be collected and evaluated in a systematic review (*Box 2 Recommendation 13*). Triangulation procedures could then be applied based on the study size, methodological quality and MCID estimates. In the end, one should consider whether a single estimate, multiple estimates or a dynamic range for the MCID evaluation best serves the requirements of the instrument under investigation.

8.5 Methodological considerations

8.5.1 Background

Data analysed in this thesis were derived from a single systematic review and meta-analysis and two clinical trials. The review (*Chapter 3*) was conducted independently by two researchers, using a predefined protocol in a structured way to explore and evaluate existing literature on the MCIDs of health status tools commonly used for COPD. It provided an overview of the methodological quality and MCID estimates and used triangulation procedures to provide a comprehensive analysis. Study 1, the *routine inspiratory muscle training within COPD rehabilitation (RIMTCORE) trial*, was then conducted in which data were included from a randomised controlled clinical trial (RCT) on the effects of IMT as add-on therapy to a 3-week PR programme for patients with COPD at the Klinik Bad Reichenhall, Germany [70]. Finally, study 2, the *MCID Study*, included patients with COPD receiving routine medical care in both primary and secondary care in the Netherlands. These patients only received care according to the Dutch COPD treatment guidelines for RCP without additional management.

8.5.2 Strengths of the study methods

The systematic review performed in *Chapter 3* offered the first structured analysis and triangulation of the literature on the MCIDs of 12 health status tools commonly used for patients with COPD [43]. A predefined and published protocol in PROSPERO was used for manuscript inclusion, analysis and assessments, which were done independently by two researchers. An extensive tool was applied to evaluate the methodological quality and risk of bias of the included studies and their MCID estimates, which comprised existing and previously used checklists for the assessment of scientific publications. Triangulation was applied for three health status instruments by incorporating elements of study size, quality assessment, MCID methods and MCID estimates. A concise overview of all results was then presented and published.

Both clinical trials also had predefined protocols that were evaluated for appropriateness by medical ethics committees. The trials included a variety of patients with COPD from two different study settings: an intervention setting (PR) and a routine medical care setting

(RCP). Data from various follow-up periods were integrated for a total duration of 12 months, and both trials had large samples at baseline (study 1 n = 451; study 2 n = 207), with loss to follow-up percentages of 68.5% for study 1 and 85.6% for study 2 due to active follow-up by the researchers (*Chapter 6* [75]). The total number of change measurements was also large, resulting in nearly 2300 scores for both samples (*Chapter 7* [80]). Study procedures were punctually followed and audited internally. Study 1 included patients with COPD GOLD grades II–IV in a strong dataset over a 3-week PR intervention. There was reasonable balance among the GOLD grades and among the baseline groups. However, slightly more males were included, and most patients were GOLD grade II or III based on spirometry. Patients were randomised to either IMT or sham-IMT by predefined randomisation lists and included via their treating physician at the clinic. Study 2 included patients with COPD GOLD grades I–IV, but the cohort had slightly more males and more patients with GOLD grades II and III disease. In this study, patients received RCP from their physician or general practitioner, and no randomisation was required.

Concerning the applied MCID methodology in the various chapters of this thesis [43, 69, 74-75, 80], it should be noted that multiple anchor- and distribution-based techniques were applied simultaneously for the three main COPD health status tools. Most previous MCID publications have focused only on a single instrument and have used only a limited number of techniques. This thesis combined and standardised the context for these frequently used health status tools for COPD to provide equal test situations. Hence, much of the resulting content was novel, providing unique data that has not previously been evaluated critically. No prior research has investigated the impact of recall bias on the MCIDs of health status instruments for COPD over multiple follow-up periods using a unique test of significance by means of dependency-adjusted confidence intervals (*Chapter 5* [74]). Equally, none have investigated the MCIDs for both deterioration and improvement at the same time (*Chapter 6* [75]). Last, but not least, no prior research has considered the various dynamic factors that affect the MCIDs of health status tools for COPD, incorporating patient-, method- and context-related aspects (*Chapter 7* [80]).

8.5.3 Limitations of the study methods

Certain limitations of the included studies should be noted. For example, many of the procedures for evaluating and determining the MCIDs of health status instruments were executed for the first time, and as such, no standard processes existed. Similarly, there were no procedures to investigate the impact of dynamic factors on the MCID estimates. This included the following elements of the thesis [43, 69, 74-75, 80]:

- the evaluation tool used for the risk of bias and quality assessment, as well as the triangulation procedures in the systematic review (*Chapter 3);*

- the application of the various anchor- and distribution-based techniques throughout the thesis;
- the evaluation method to assess the impact of the recall period upon the MCID estimates (*Chapter 5*);
- the subgroup analysis for baseline-, context- and patient-related factors (*Chapter 7*); and
- the development and application of dynamic linear multiple regression models (*Chapter 7*).

The lack of standardised procedures for evaluating the MCID became evident in the systematic review (*Chapter 3* [43]). Here, one was forced to use a new purpose-built tool for the quality and risk of bias assessments, and although this was based on existing checklists, the final instrument itself was not validated. The absence of an established checklist for evaluating studies of an instrument's MCID necessitated this decision. By integrating elements from validated checklists, one can have some confidence in its validity. Coupled with this, the triangulation procedures applied in the systematic review were specifically developed for this research, because no current procedures existed. By including a variety of aspects, such as study size, quality, and MCID methods and estimates, triangulation was considered valid. Indeed, it was unfortunate that triangulation could not be performed for all included health status instruments due to limited number of studies or their heterogeneity. In addition, some studies that had claimed to investigate the MCID of instruments, were excluded from the review for failing to meet the inclusion criteria. This is a pity too, because it would have been worthwhile to evaluate these critically. Another important point here is that the systematic review included the article presented in *Chapter 4* of this thesis [69] because of the timing of publication and the required search update. It was one of the largest studies included, and this may have affected the triangulation results.

Regarding the RIMTCORE trial (study 1), which compared IMT with placebo (sham IMT) [70], no differences were observed between the intervention arms in the MCID analyses of this thesis. All data were analysed collectively with those patients meeting the inclusion requirements for MCID analysis, but the analyses were performed retrospectively and without corrections between the intervention arms. This RCT design may have affected the findings in this thesis. By comparison, the MCID study (study 2) had a simpler design, requiring patients to assess their health statuses by questionnaire at various times at home. Patients were included by telephone, mail, general practitioner or pulmonary physician. No formal face-to-face evaluation of the patient was made by the researcher. The administration of questionnaires at home could also have biased the follow-up results.

It is important to note that not all chapters in this thesis included data from both studies, mainly due to the timing of the various publications. Incorporating data from both settings in each chapter would have made the analyses stronger, especially in *Chapters 4 and 5* [69, 74]. While comparing the included RIMTCORE trial and MCID study, significant baseline differences were observed in terms of age, spirometry and health status results, with no corrections made for the exploratory nature of this thesis. Furthermore, only the MCID study included patients with COPD GOLD grade I, making comparisons between the datasets more difficult. In addition, the RIMTCORE trial included significantly more patients than the MCID study, giving a greater weighting to the patients receiving PR in the combined MCID analyses (*Chapter 7* [80]). Moreover, more follow-up periods and measurement moments were used in the RIMTCORE trial, again providing its data with greater weight. Some benefit would also have been gained from incorporating data from other interventions, thereby allowing discussion of whether the results obtained in this thesis were valid for PR only or for interventions in general.

In the various MCID analyses of *Chapters 4–7* [69, 74-75, 80], many different anchor- and distribution-based methods were used. However, not all chapters included all methods simultaneously in an effort to preserve the overview and prevent the reader from being overburdened by the sheer number of MCID estimates. Given that *Chapter 4* demonstrated that there was good agreement between several techniques, one may argue that not all methods would be required. However, it was unfortunate that a limited number of patients with deterioration, or with moderate or large changes, were included for the anchor-based methods. This resulted in small subgroups with larger confidence intervals. The anchor-based mean change method used this kind of subgroup analysis based on the GRC categorisation. Although the overall sample sizes were large in both studies, the numbers of patients in each subgroup were limited, reducing the power.

Another major limitation of the anchor-based methods applied here was that some correlations between the selected anchors and questionnaires under review were insufficiently strong. This was especially true directly after PR intervention. Furthermore, *Chapter* 6 [75] indicated that there was a response shift, with stronger correlations between the GRC and follow-up scores than with the change scores. No corrections were made for this finding, which could limit the conclusions. It is unknown how to correct for this in MCID research. Next, it was noted that the design of the GRCs could affect the MCID estimates (*Chapter 5* [74]). Preferably, a greater number of different GRCs would have been applied during the various follow-up periods in each setting, but just two anchor questions were applied and only at the 12-month follow-up assessment for the PR intervention. Moreover, it is important to state that the assessment of change with the GRC was performed by the patient, and that there was no formal clinical assessment

by the physician to confirm the self-reported data. The costs of treatment were also ignored when reaching the MCID. Finally, concerning the anchor-based methods, it is worth noting that exacerbations were used as clinical events for criterion-referencing. Although good agreement was observed between the patient- and criterion-referencing (*Chapter 4* [69]), exacerbations may not necessarily represent a *minimal event*. It would also have been interesting to execute the criterion-referencing approach during RCP in study 2.

A limitation of the conclusions presented in *Chapter 7* of this thesis [80] was that a complex interaction existed between the various factors affecting the MCID. Indeed, the models and interaction terms only explained 20% to 40% of the MCID estimates in the final analyses, and not all the observed trends were significant. Therefore, due to the inherent complexity, this thesis could not provide a comprehensive answer on the scope of the MCID dynamics. A regression to the mean phenomenon could explain this finding, but despite being quantified, it was not further integrated in the results. It is unknown how to incorporate this in MCID research.

8.6 Clinical implications and suggestions for future research

8.6.1 Suggestions for researchers and physicians concerning MCID research in general

The current thesis demonstrated that the MCID is a complex and dynamic parameter for which not all details could be fully explained or clarified. For general MCID measurement, a set of evidence-based recommendations was developed to guide researchers on how to determine an instrument's MCID with greater scientific rigor (*Box 2*). Future instruments will benefit from incorporating these guidelines in MCID determination processes. The recommendations can be applied to many different (physiological) tests and tools, not specifically those limited to health status, having potential utility with any measurement outcome used to interpret treatment effects. It is highly recommended that these guidelines are integrated into future MCID research. Moreover, it is important to evaluate the level to which they are applicable and whether they hold firm for instruments used with chronic diseases other than COPD. It is especially interesting to understand the dynamics of the MCID model better. This will help with the interpretation of more tailored, clustered or even individual change scores that are of more use to physicians in clinical practice. If the recommendations from this thesis do not hold firm, they should be adapted to each setting, which may be facilitated by a consensus meeting of experts.

An important consideration that arose from our findings and recommendations for MCIDs was that many existing questionnaires, PROs and physiological parameters may benefit from researchers re-evaluating their currently applied MCIDs and techniques used to obtain them. Gaps in the body of evidence should be noted, and where possible, filled with evidence from new studies based on the recommendations set out in *Box 2*. Clinical trials should incorporate relevant anchors to be able to determine anchor- and distribution-based MCIDs on a large scale with evidence from multiple settings and populations. There can be no excuse for the continued use of insufficiently founded MCIDs when interpreting treatment outcomes in the future.

8.6.2 Suggestions for researchers and physicians concerning COPD research

The current thesis should by no means be considered the end of the evaluation process for the MCIDs of health status tools for patients with COPD. Many of the outcomes of this thesis were based on first time thorough research into the dynamics of this parameter. The results need to be either confirmed or refuted in clinical replication studies, potentially building on the limited number and methodological quality of existing research. Incorporating relevant anchors in clinical trials from multiple settings and samples would enable to investigate MCIDs in COPD health status at a large scale with anchor- and distribution-based methods. This is especially true when defining the MCIDs for deterioration, moderate clinically important changes and large clinically important changes. It is also relevant to discover the full dynamics of the MCIDs for health status tools for COPD. Despite having now taken the first steps towards validating these thresholds, more evidence should be collected to determine whether the MCIDs for the domain scores of the CCQ and SGRQ are indeed similar to the overall patterns observed for the MCID.

There is also a need to confirm whether the differences observed between the MCIDs for the PR and RCP settings are applicable to other interventions too, not least because these differences were not confirmed by the systematic review in *Chapter 3* [43]. Based on the outcomes, of this systematic review, it could be argued that MCIDs for different interventions may expected to be similar and not just context-specific to PR only. One may also wonder why MCIDs for various interventions would need to differ. Should thresholds for minimal clinically important change be lower if the effect of outcomes of a specific intervention are expected to be less than in PR? It would make more sense to argue that patient characteristics and baseline severity would have more impact on the value of the required MCID threshold than a specific setting or intervention. This would be expected because patient characteristics and baseline severity seem highly correlated with the type of a specific context or setting.

An important implication of this thesis is that current COPD treatment guidelines may be grounded on MCIDs of questionable robustness. The MCIDs of the SGRQ (4 points) and the CRQ (0.5 points) have played especially large roles in studies of pharmaceutical and other medical interventions. However, these MCID estimates appear to have been inadequately researched, with evidence that for the SGRQ one should use higher thresholds than are appropriate, especially when used to assess interventions or change directly after therapy. As such, any research based on its MCID over recent decades may have overestimated the clinical trial effects. The currently used MCIDs should be higher for both improvement and deterioration, raising two important questions: (1) Would this mean that most of the scientific evidence, currently integrated in treatment guidelines, is invalid and should not have been approved for patients? (2) Does the basis for certain therapies and treatment regimens remain solid? Patients in clinical practice seem to respond to treatments based on the traditional health status MCIDs; if not, they may have stopped using a given medication or have stopped participation in an intervention. Future research should systematically review the validity of current evidence and thresholds for MCIDs. Consensus should be reached between COPD experts on how to handle this and on the extent to which treatment guidelines are affected.

A more dynamic framework for evaluating treatment effects by the MCID was proposed for the CAT, CCQ and SGRQ (*Table 2, Figure 4*). This could serve not only as a guideline for use in COPD research, but also as an example of appropriate MCID development in other chronic disease research. The dynamic framework requires more flexibility from scientists in research and physicians in clinical practice, but benefits from integrating both the point estimate and its 95%CI as an outcome of a clinical trial. This modified approach also benefits from (more individual) assessment by shifting up or down categories based on differences in setting, follow-up period, baseline severity (including extreme groupings) and patient characteristics. One should, however, be cautious for an overly flexible interpretation of treatment effects. It is important to discuss the criteria that need to be considered in this dynamic framework, and further research should seek to confirm or refute the purported benefits of the framework. Validating the proposed dynamic MCID framework in scientific trials and clinical practice is highly recommended to confirm or falsify this thesis' findings.

Given that the MCIDs for the traditional SGRQ and CRQ health status tools were less firm than those for the CAT and CCQ, it would be argued that these latter instruments should be used in future research and clinical practice; and it is promising to see that trials are increasingly doing so when evaluating outcomes. Based on the findings of this thesis, their MCIDs for deterioration could now be applied, and due to the differing MCID dynamics, it may even be appropriate to suggest that clinical trials should integrate more than one

health status tool when evaluating outcomes. Irrespective of the route chosen, it appears that the currently accepted thresholds for clinically relevant change measured by the SGRQ and CRQ need to be applied more carefully. Moreover, there is definitely no justification for using the MCIDs of other health status instruments for COPD, because existing evidence is of poor quality and/or quantity (*Chapter 3* [43]).

A final important implication is that frequently applied MCIDs for physiological parameters may now need to be re-evaluated according to this thesis' recommendations (*Box 2*). These include the forced expiratory volume in one second (FEV_1), six minute walking distance (6MWD) and exacerbations. Currently, these physiological MCIDs are also used as obligatory primary outcome parameters when interpreting observed changes in scientific trials and clinical practice, but questions may now exist over how their MCIDs were determined. This was beyond the scope of this thesis and must be investigated further.

8.6.3 Suggestions for physicians in clinical practice

In clinical practice, physicians require tests and tools that can help to interpret the effects of therapy in individual patients. Today, PROs are integrated into many consultations along with history taking and physical examination, but physicians require guidance on how to interpret the observed changes on PROs (e.g., whether an observed difference in an item score represents a real change or just random variation). Unfortunately, there is no straight forward answer for the physician due to the large individual variation of the patients. MCIDs are defined at the group level. The proposed dynamic framework for MCID interpretation (*Table 2, Figure 4*) offers a reasonable starting point to aid individual interpretation of changes observed, also for clinical practice. The interpretation using this framework could be tailored to specific patient characteristics, contexts and/or baseline (health status) severities by requiring a more or less firm categorisation of the changes observed. Personality, comorbidities and spirometry severity could perhaps be incorporated too. Validating this framework at the individual patient level would be highly recommended. Based on a recent study evaluating expert opinions on the MCID of the SGRQ, thresholds for minimal clinically important change would even need to exceed 10 points for individual patients [121]. This would suggest that for an individual interpretation of change scores, the classification of patients on the dynamic MCID framework would need to be more strict.

It would perhaps even be worthwhile to develop the dynamic aspects of the MCID into a formula (score) that facilitates the change assessments in both primary and secondary care. To develop a more tailored MCID formula, one would need to confirm or refute the trends observed in *Chapter 7* of this thesis [80]. These concerned the impact of the various patient-, disease- and context-related factors on the MCID. Ideally a

large database of clinical trial data should be combined with clinical event data and patient- or clinician-assessed GRC anchors. In this way, MCIDs can be determined and developed into formulas for tailored evaluation of individual change scores. This would need to originate from international collaboration and should include data on patient characteristics, setting, intervention, clinical events and measurement period, as well as a rating of the clinical importance of the observed change.

Finally, it is important for physicians in clinical practice to remain aware of how MCIDs play a role in interpreting treatment outcomes and approving new therapies. They need to be alerted to this, especially when reading and evaluating national and international treatment guidelines. Understanding the dynamics of the MCID concept is therefore crucial if physicians are to remain critical when interpreting literature and guidelines.

8.7 Final conclusions

The MCID is a complex parameter that is key to the interpretation of treatment outcomes. Although its dynamics have not been fully explored, there are few alternatives for evaluating and interpreting therapy effects. As such, the MCID remains of vital importance, but it should also be treated with caution. Current practices use MCIDs as static and fixed thresholds that lack the flexibility to deal with the existing limitations. Therefore, a framework was suggested in this thesis to facilitate the dynamic and flexible interpretation of observed changes in patients with COPD. Coupled with this, general recommendations were presented to guide a more evidence-based approach to determining an instrument's MCID. Developing an instrument's MCID requires that a structural and uniform approach be adopted, and one must certainly endeavour to protect against this process continuing to be taken for granted. Failure to do so could result in false MCIDs that cause overestimations or underestimations of clinical and scientific outcomes.

8.8 References

[1] Jones, PW. Health status measurement in chronic obstructive pulmonary disease. *Thorax* 2001; 56(11): 880-887.

[2] Curtis, JR and Patrick, DL. The assessment of health status among patients with COPD. *European Respiratory Journal Supplement* 2003; 41: 36s-45s.

[3] Reardon, JZ, Lareau, SC and ZuWallack, R. Functional Status and Quality of Life in Chronic Obstructive Pulmonary Disease. *The American Journal of Medicine* 2006; 119(10 Suppl 1): S32-37.

[4] Crosby, RD, Kolotkin, RL and Williams, GR. Defining clinically meaningful change in health-related quality of life. *Journal of Clinical Epidemiology* 2003; 56(5): 395-407.

[5] Patrick, DL and Bergner, M. Measurement of Health Status in the 1990s. *Annual Review of Public Health* 1990; 11: 165-183.

[6] Fitzpatrick, R, Davey, C, Buxton, MJ *et al*. Evaluating patient-based outcome measures for use in clinical trials. *Health Technology Assessment* 1998; 2(14): 1-74.

[7] Bergner, M. Quality of Life, Health Status and Clinical Research. *Medical Care* 1989; 27(3 Suppl): s148-156.

[8] Guyatt, GH, Feeny, DH and Patrick, DL. Measuring health-related quality of life. *Annals of Internal Medicine* 1993; 118(8): 622-629.

[9] Jones, PW. Issues Concerning Health-Related Quality of Life in COPD. *Chest* 1995; 107(5 Suppl): 187S-193S.

[10] Jones, PW. Health status and the spiral of decline. *COPD: Journal of Chronic Obstructive Pulmonary Disease* 2009; 6(1): 59-63.

[11] Jones, PW. Health Status: What Does It Mean for Payers and Patients? *Proceedings of the American Thoracic Society* 2006; 3(3): 222-226.

[12] US Department of Health and Human Services FDA Center for Drug Evaluation and Research. Guidance for industry: patient-reported outcome measures: use in medical product development to support labelling claims: draft guidance. *Health and Quality of Life Outcomes* 2006; 4: 79.

[13] Jones, P, Miravittles, M, van der Molen, T *et al*. Beyond FEV1 in COPD: a review of patient-reported outcomes and their measurement. *International Journal of COPD* 2012; 7: 697-709.

[14] Global initiative for Chronic Obstructive Lung Disease (GOLD). Global Strategy for the Diagnosis, Management and Prevention of COPD (2017 Report). 2017. Available at http://www.goldcopd.org/gold-2017-global-strategy-diagnosis-management-prevention-copd/. Accessed 9th of July 2018.

[15] Barnett, M. Chronic obstructive pulmonary disease: a phenomenological study of patients' experiences. *Journal of Clinical Nursing* 2005; 14(7): 805-812.

[16] Rennard, SI and Drummond, MB. Early chronic obstructive pulmonary disease: definition, assessment, and prevention. *Lancet* 2015; 385(9979): 1778-1788.

[17] Westwood, M, Bourbeau, J, Jones, PW *et al*. Relationship between FEV1 change and patient-reported outcomes in randomized trials of inhaled bronchodilators for stable COPD: a systematic review. *Respiratory Research* 2011; 12: 40.

[18] Tsiligianni, I, Kocks, J, Tzanakis, N *et al*. Factors that influence disease-specific quality of life or health status in patients with COPD: a review and meta-analysis of Pearson correlations. *Primary Care Respiratory Journal* 2011; 20(3); 257-268.

[19] Martin, AL, Marvel, J, Fahrbach, K *et al*. The association of lung function and St. George's respiratory questionnaire with exacerbations in COPD: a systematic literature review and regression analysis. *Respiratory Research* 2016; 17: 40.

[20] Bergner, M and Rothman, ML. Health Status Measures: An Overview and Guide for Selection. *Annual Review of Public Health* 1987; 8: 191-210.

[21] Celli, BR, MacNee, W, ATS/ERS Task Force. Standards for the diagnosis and treatment of patients with COPD: a summary of the ATS/ERS position paper. *European Respiratory Journal* 2004; 23(6): 932-946.

[22] Hogg, JC. Pathophysiology of airflow limitation in chronic obstructive pulmonary disease. *Lancet* 2004; 364(9435): 709-721.

[23] Murray, CJL *et al*. Global, regional, and national age-sex specific mortality for 264 causes of death, 1980–2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet* 2017; 390(10100): 1151-1210.

[24] Murray, CJL, Vos, T, Lozano, R *et al*. Disability-adjusted Life Years (DALYs) for 291 diseases and injuries in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet* 2012; 380(9859): 2197-2223.

[25] Vos, T, Flaxman, AD, Naghavi, M *et al*. Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet* 2012; 380(9859): 2163-2196.

[26] Negewo, A, Gibson, PG and McDonald, VM. COPD and its comorbidities: Impact, measurement and mechanisms. *Respirology* 2015; 20(8): 1160-1171.

[27] Corlateanu, A, Covantev, S, Mathioudakis, AG *et al*. Prevalence and burden of comorbidities in Chronic Obstructive Pulmonary Disease. *Respiratory Investigation* 2016; 54(6): 387-396.

[28] Divo, MJ, Casanova, C, Marin, JM *et al*. Chronic obstructive pulmonary disease comorbidities network. *European Respiratory Journal* 2015; 46(3): 640-650.

[29] Miravitlles, M, Vogelmeier, C, Roche, N *et al*. A review of national guidelines for management of COPD in Europe. *European Respiratory Journal* 2016; 47(2): 625-637.

[30] Cave, J, Atkinson, L, Tsiligianni, IG *et al*. Assessment of COPD wellness tools for use in primary care: an IPCRG initiative. *International Journal of COPD* 2012; 7: 447-456.

[31] Jones, PW, Quirck, FH and Baveystock, CM. The St George's Respiratory Questionnaire. *Respiratory Medicine* 1991; 85 (Suppl B): 25-31.

[32] Jones, PW, Quirck, FH, Baveystock, M *et al*. A Self-complete Measure of Health Status for Chronic Airflow Limitation: The St. George's Respiratory Questionnaire. *American Review of Respiratory Disease* 1992; 145(6): 1321-1327.

[33] Jones, PW, Harding, G, Berry, P *et al*. Development and first validation of the COPD Assessment Test. *European Respiratory Journal* 2009; 34(3): 648-654.

[34] Van der Molen, T, Willemse, BWM, Schokker, S *et al*. Development, validity and responsiveness of the Clinical COPD Questionnaire. *Health and Quality of Life Outcomes* 2003, 1: 13.

[35] Davidson, M and Keating, J. Patient-reported outcome measures (PROMS): how should I interpret reports of measurement properties? A practical guide for clinicians and researchers who are not biostatisticians. *British Journal of Sports Medicine* 2014; 48(9): 792-796.

[36] Jones, PW, Rennard, S, Tabberer, M *et al*. Interpreting patient-reported outcomes from clinical trials in COPD: a discussion. *International Journal of COPD* 2016; 11: 3069-3078.

[37] Beaton, DE, Boers, M and Wells, GA. Many faces of the minimal clinically important difference (MCID): a literature review and directions for future research. *Current Opinion in Rheumatology* 2002; 14(2): 109-114.

[38] Jaeschke, R, Singer, J and Guyatt, GH. Measurement of health status. Ascertaining the Minimal Clinically Important Difference. *Controlled Clinical Trials* 1989; 10(4): 407-415.

[39] Sloan, JA. Assessing the Minimally Clinically Significant Difference: Scientific Considerations, Challenges and Solutions. *COPD: Journal of Chronic Obstructive Pulmonary Disease* 2005; 2(1): 57-62.

[40] Jones, PW. St. George's respiratory questionnaire: MCID. *COPD* 2005; 2(1): 75-79.

[41] Zhou, Z, Zhou, A, Zhao, Y *et al*. Evaluating the Clinical COPD Questionnaire: A systematic review. *Respirology* 2017; 22(2): 251-262.

[42] Gupta, N, Pinto, LM, Morogan, A *et al*. The COPD assessment test: a systematic review. *European Respiratory Journal* 2014; 44(4): 873-884.

[43] Alma, H, de Jong, C, Tsiligianni, I *et al*. Clinically relevant differences in COPD health status: systematic review and triangulation. *European Respiratory Journal* 2018; 52(3): 1800412.

[44] Copay, AG, Subach, BR, Glassman, SD *et al*. Understanding the minimum clinically important difference: a review of concepts and methods. *The Spine Journal* 2007; 7(5): 541-546.

[45] Make, B. How Can We Assess Outcomes of Clinical Trials: The MCID Approach. *COPD: Journal of Chronic Obstructive Pulmonary Disease* 2007; 4(3): 191-194.

[46] Make, B, Casaburi, R and Leidy, NK. Interpreting Results from Clinical Trials: Understanding Minimal Clinically Important Differences in COPD Outcomes. *COPD: Journal of Chronic Obstructive Pulmonary Disease* 2005; 2(1): 1-5.

[47] Revicki, D, Hays, RD, Cella, D *et al*. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *Journal of Clinical Epidemiology* 2008; 61(2): 102-109.

[48] Man-Son Hing, M, Laupacis, A, O'Rourke, K *et al*. Determination of the Clinical Importance of Study Results: A Review. *Journal of General Internal Medicine* 2002; 17(6): 469-476.

[49] Leidy, NK and Wyrwich, KW. Bridging the Gap: Using Triangulation Methodology to Estimate Minimal Clinically Important Differences (MCIDs). *COPD: Journal of Chronic Obstructive Pulmonary Disease* 2005; 2(1): 157-165.

[50] Cook, CE. Clinimetrics Corner: The Minimal Clinically Important Change Score (MCID): A Necessary Pretense. *Journal of Manual & Manipulative Therapy* 2008; 16(4): 82-83.

[51] Hays, RD, Farivar, SS and Liu, H. Approaches and Recommendations for Estimating Minimally Important Differences for Health-Related Quality of Life Measures. *COPD: Journal of Chronic Obstructive Pulmonary Disease* 2005; 2(1): 63-67.

[52] Angst, F, Aeschlimann, A and Angst, J. The minimal clinically important difference raised the significance of outcome effects above the statistical level, with methodological implications for future studies. *Journal of Clinical Epidemiology* 2017; 82: 128-136.

[53] Hays, RD and Woolley, JM. The Concept of Clinically Meaningful Difference in Health-Related Quality of Life Research. How Meaningful is it? *Pharmacoeconomics* 2000; 18(5): 419-423.

[54] Dworkin, RH, Turk, DC, Wyrwich, KW *et al*. Consensus Statement Interpreting the Clinical Importance of Treatment Outcomes in Chronic Pain Clinical Trials: IMMPACT Recommendations. *The Journal of Pain* 2008; 9(2): 105-121.

[55] Jones, PW. Interpreting thresholds for a clinically significant change in health status in asthma and COPD. *European Respiratory Journal* 2002; 19(3): 398-404.

[56] Norman, GR. The Relation Between the Minimally Important Difference and Patient Benefit. *COPD: Journal of Chronic Obstructive Pulmonary Disease* 2005; 2(1): 69-73.

[57] Norman, GR, Sloan, JA and Wyrwich, KW. Interpretation of Changes in Health-related Quality of Life. The Remarkable Universality of Half a Standard Deviation. *Medical Care* 2003; 41(5): 582-592.

[58] Ostelo, RWJG, Deyo, RA, Stratford, P *et al*. Interpreting Change Scores for Pain and Functional Status in Low Back Pain. Towards International Consensus Regarding Minimal Important Change. *Spine* 2008; 33(1): 90-94.

[59] Revicki, DA, Cella, D, Hays, RD *et al*. Responsiveness and minimal important differences for patient reported outcomes. *Health and Quality of Life Outcomes* 2006; 4: 70.

[60] Turner, D, Schünemann, HJ, Griffith, LE *et al*. The minimal detectable change cannot reliably replace the minimal important difference. *Journal of Clinical Epidemiology* 2010; 63(1): 28-36.

[61] Wyrwich, KW and Wolinsky, FD. Identifying meaningful intra-individual change standards for health-related quality of life measures. *Journal of Evaluation in Clinical Practice* 2000; 6(1): 39-49.

[62] Wells, G, Beaton, D, Shea, B *et al*. Minimal Clinically Important Differences: Review of Methods. *The Journal of Rheumatology* 2001; 28(2): 406-412.

[63] Wyrwich, KW, Fihn, SD, Tierney, WM *et al*. Clinically Important Changes in Health-related Quality of Life for Patients with Chronic Obstructive Pulmonary Disease. An Expert Consensus Panel Report. *Journal of General Internal Medicine* 2003; 18(3): 196-202.

[64] Middel, B. Statistical significant change versus relevant or important change in (quasi) experimental design: Some conceptual and methodological problems in estimating magnitude of intervention-related change in health services research. *International Journal of Integrated Care* 2002; 2: e15.

[65] Redelmeier, DA, Guyatt, GH and Goldstein, RS. Assessing the Minimal Important Difference in Symptoms: A Comparison of Two Techniques. *Journal of Clinical Epidemiology* 1996; 49(11): 1215-1219.

[66] Wyrwich, KW, Nienaber, NA, Tierney, WM *et al*. Linking Clinical Relevance and Statistical Significance in Evaluating Intra-Individual Changes in Health-Related Quality of Life. *Medical Care* 1999; 37(5): 469-478.

[67] Wyrwich, KW, Tierney, WM and Wolinsky, FD. Further Evidence Supporting an SEM-Based Criterion for Identifying Meaningful Intra-Individual Changes in Health-Related Quality of Life. *Journal of Clinical Epidemiology* 1999; 52(9): 861-873.

[68] Engel, L, Beaton, DE and Touma, Z. Minimal Clinically Important Difference: a Review of Outcome Measure Score Interpretation. *Rheumatic Disease Clinics North America* 2018; 44(2): 177-188.

[69] Alma, H, de Jong, C, Jelusic, D *et al*. T. Health status instruments for patients with COPD in pulmonary rehabilitation: defining a minimal clinically important difference. *npj Primary Care Respiratory Medicine* 2016; 26: 16041.

[70] Schultz, K, Jelusic, D, Wittmann, M *et al.* Inspiratory muscle training does not improve clinical outcomes in 3-week COPD rehabilitation: results from a randomised controlled trial. *European Respiratory Journal* 2018; 51(1): 1702000.

[71] Kamper, SJ, Maher, C and Mackay, G. Global Rating of Change Scales: A Review of Strengths and Weaknesses and Considerations for Design. *Journal of Manual & Manipulative Therapy* 2009; 17(3): 163-170.

[72] Grovle, L, Haugen, AJ, Hasvik, E *et al*. Patients ratings of global perceived change during 2 years were strongly influenced by the current health status. *Journal of Clinical Epidemiology* 2014; 67(5): 508-515.

[73] Jaeschke, RJ, Guyatt, GH, Keller, J *et al*. Interpreting Changes in Quality of Life Score in N of 1 Randomized Trials. *Controlled Clinical Trials* 1991; 12(4 Suppl): 226S-233S.

[74] Alma, H, de Jong, C, Jelusic, D *et al*. Assessing health status over time: impact of recall period and anchor question on the minimal clinically important difference of copd health status tools. *Health and Quality of Life Outcomes* 2018; 16(1): 130.

[75] Alma, HJ, de Jong, C, Jelusic, D *et al.* Thresholds for clinically important deterioration versus improvement in COPD health status: results from a randomized controlled trial in pulmonary rehabilitation and an observational study during routine clinical practice. *BMJ Open* 2019; 9(6): e025776.

[76] Hajiro, TH and Nishimura, K. Minimal clinically significant difference in health status: the thorny path of health status measures? *European Respiratory Journal* 2002; 19(3): 390-391.

[77] Jones, PW, Beeh, KM, Chapman, KR *et al*. Minimal Clinically Important Differences in Pharmacological Trials. *American Journal of Respiratory Critical Care Medicine* 2014; 189(3): 250-255.

[78] Troosters, T. How important is a minimal difference? *European Respiratory Journal* 2011; 37(4): 755-756.

[79] Chapman, KR, Bergeron, C, Bhutani, M *et al*. Do We Know the Minimal Clinically Important Difference (MCID) for COPD Exacerbations? *COPD: Journal of Chronic Obstructive Pulmonary Disease* 2013; 10(2): 243-249.

[80] Alma, H, de Jong, C, Jelusic, D *et al*. Baseline health status and setting impacted minimal clinically important differences in COPD: an exploratory study. *Journal of Clinical Epidemiology* 2019; 116: 49-61.

[81] Gottfried, MH, Kerwin, EM, Lawrence, D *et al*. Efficacy of indicaterol 75µg once-daily on dyspnea and health status results of two double-blind, placebo controlled 12-week studies. *Chronic Obstructive Pulmonary Disease* 2012; 9(6): 629-636.

[82] Jones, PW, Price, D and van der Molen, T. Role of clinical questionnaires in optimizing everyday care of chronic obstructive pulmonary disease. *International Journal of COPD* 2011; 6: 289-296.

[83] Tsiligianni, IG, van der Molen, T, Moraitaki, D *et al*. Assessing health status in COPD. A head-to-head comparison between the COPD assessment test (CAT) and the clinical COPD questionnaire (CCQ). *BMC Pulmonary Medicine* 2012, 12: 20.

[84] Sundh, J, Ställberg, B, Lisspers, K *et al*. Comparison of the COPD Assessment Test (CAT) and the Clinical COPD Questionnaire (CCQ) in a Clinical Population. *COPD* 2016; 13(1): 57-65.

[85] Ringbaek, T, Martinez, G and Lange, P. A Comparison of the Assessment of Quality of Life with CAT, CCQ, and SGRQ in COPD Patients Participating in Pulmonary Rehabilitation. *COPD* 2012; 9(1): 12-15.

[86] Guyatt, GH, Berman, LB, Townsend, M *et al*. A measure of quality of life for clinical trials in chronic lung disease. *Thorax* 1987; 42(10): 773-778.

[87] Schünemann, HJ, Griffith, L, Jaeschke, R *et al*. Evaluation of the minimal important difference for the feeling thermometer and the St. George's Respiratory Questionnaire in patients with chronic airflow obstruction. *Journal of Clinical Epidemiology* 2003; 56(12): 1170-1176.

[88] Guyatt, GH, Norman, GR, Juniper, EF and Griffith, LE. A critical look at transition ratings. *Journal of Clinical Epidemiology* 2002; 55(9): 900-908.

[89] Stull, DE, Klinke Leidy, N, Parasuraman, B *et al*. Optimal recall periods for patient-reported outcomes: challenges and potential solutions. *Current Medical Research Opinion* 2009; 25(4): 929-942.

[90] Kjellson, G, Clarke, P and Gerdtham, UG. Forgetting to remember or remembering to forget: A study of the recall period length in health care survey questions. *Journal of Health Economics* 2014; 35: 34-46.

[91] Schmitt, JS and Abbott, JH. Global ratings of change do not accurately reflect functional change over time in clinical practice. *Journal of Orthopaedics and Sports Physical Therapy* 2015; 45(2): 106-111.

[92] Schmitt, JS and Abbott, JH. Patient global ratings of change did not adequately reflect change over time: a clinical cohort study. *Physical Therapy* 2014; 94(4): 534-542.

[93] Schmitt, J and Di Fabio, RP. The Validity of Prospective and Retrospective Global Change Criterion Measures. *Archives of Physical Medical Rehabilitation* 2005; 86(12): 2270-2276.

[94] Kamper, SJ, Ostelo, RW, Knol, DL *et al.* Global Perceived Effect scales provided reliable assessments of health transition in people with musculoskeletal disorders, but ratings are strongly influenced by current status. *Journal of Clinical Epidemiology* 2010; 63(7): 760-766.

[95] Norman, GR, Stratford, P and Regehr, G. Methodological problems in the retrospective computation of responsiveness to change: the lesson of Cronbach. *Journal of Clinical Epidemiology* 1997; 50(8): 869-879.

[96] Herrmann, D. Reporting current, past, and changed health status. What we know about distortion. *Medical Care* 1995; 33(4 Suppl): AS89-94.

[97] Meek, PM, Lareau, SC and Anderson, D. Memory for symptoms in COPD patients: how accurate are their reports? *European Respiratory Journal* 2001; 18(3): 474-481.

[98] Sundh, J, Montgomery, S, Hasselgren, M *et al*. Change in health status in COPD: a seven-year follow-up cohort study. *npj Primary Care Respiratory Medicine* 2016; 26: 16073.

[99] Nagai, K, Makita, H, Suzuki, M *et al*. Differential changes in quality of life components over 5 years in chronic obstructive pulmonary disease patients. *International Journal of COPD* 2015; 10: 745-757.

[100] Singh, D, D'Urzo, AD, Chuecos, F *et al.* Reduction in clinically important deterioration in chronic obstructive pulmonary disease with aclidinium/formoterol. *Respiratory Research* 2017; 18(1): 106.

[101] Raman, S, Ding, K, Chow, E *et al.* Minimal clinically important differences in the EORTC QLQ-BM22 and EORTC QLQ-C15-PAL modules in patients with bone metastases undergoing palliative radiotherapy. *Quality of Life Research* 2016; 25(10): 2535-2541.

[102] Maringwa, J, Quinten, C, King, M *et al.* Minimal clinically meaningful differences for the EORTC QLQ-C30 and EORTC QLQ-BN20 scales in brain cancer patients. *Annals of Oncology* 2011; 22(9): 2107-2112.

[103] Hui, D, Shamieh, O, Paiva, CE *et al.* Minimal Clinically Important Difference in the Physical, Emotional, and Total Symptom Distress Scores of the Edmonton Symptom Assessment System. *Journal of Pain and Symptom Management* 2016; 51(2): 262-269.

[104] Tuomi, L, Johansson, M, Andréll, P and Finizia, C. Interpretation of the Swedish Self Evaluation of Communication Experiences after Laryngeal cancer: Cutoff levels and minimum clinically important differences. *Head Neck* 2016; 38(5): 689-695.

[105] Rouquette, A, Blanchin, M, Sébille, V *et al.* The minimal clinically important difference determined using item response theory models: an attempt to solve the issue of the association with baseline score. *Journal of Clinical Epidemiology* 2014; 67(4): 433-440.

[106] Hui, D, Shamieh, O, Paiva, CE *et al.* Minimal clinically important differences in the Edmonton Symptom Assessment Scale in cancer patients: A prospective, multicentre study. *Cancer* 2015; 121(17): 3027-3035.

[107] de Vet, HC, Terwee, CB, Ostelo, RW *et al.* Minimal changes in health status questionnaires: distinction between minimally detectable change and minimally important change. *Health and Quality of Life Outcomes* 2006; 4: 54.

[108] Wright, JG. The Minimal Important Difference: Who's to Say What Is Important? *Journal of Clinical Epidemiology* 1996; 49(11): 1221-1222.

[109] Maltenfort, M and Diaz-Ledezma, C. Statistics In Brief: Minimum Clinically Important Difference – Availability of Reliable Estimates. *Clinical Orthopaedics and Related Research* 2017; 475(4): 933-946.

[110] Wang, YC, Hart, DL, Stratford, PW and Mioduski, JE. Baseline Dependency of Minimal Clinically Important Improvement. *Physical Therapy* 2011; 91(5): 675-688.

[111] Terwee, CB, Roorda, LD, Dekker, J *et al*. Mind the MIC: large variation among populations and methods. *Journal of Clinical Epidemiology* 2010; 63(5): 524-534.

[112] Arima, H, Carreon, LY, Glassman, SD *et al*. Age variation in the minimum clinically important difference in SRS-22r after surgical treatment for adult spinal deformity – A single institution analysis in Japan. *Journal of Orthopedic Science* 2018; 23(1): 20-25.

[113] Katz, NP, Paillard, FC and Ekman, E. Determining the clinical importance of treatment benefits for interventions for painful orthopedic conditions. *Journal of Orthopedic Surgery and Research* 2015; 10: 24.

[114] Stucki, G, Daltroy, L, Katz, JN *et al*. Interpretation of change scores in ordinal clinical scale and health status measures: the whole may not equal the sum of the parts. *Journal of Clinical Epidemiology* 1996; 49(7): 711-717.

[115] Baker, DW, Hays, RD and Brook, RH. Understanding Changes in Health Status: Is the Floor Phenomenon Merely the Last Step of the Staircase? *Medical Care* 1997; 35(1): 1-15.

[116] Stratford, PW, Binkley, JM, Riddle, DL and Guyatt, GH. Sensitivity to change of the Roland-Morris Back Pain Questionnaire: part 1. *Physical Therapy* 1998; 78(11): 1186-1196.

[117] Schuller, W, Ostelo, RW, Janssen, R and de Vet, HC. The Influence of study population and definition of improvement on the smallest detectable change and the minimal important change of the neck disability index. *Health and Quality of Life Outcomes* 2014; 12: 53.

[118] Farrar, JT, Young, JP Jr, LaMoreaux, L *et al.* Clinical importance of changes in chronic pain intensity measured in an 11-point numerical pain rating scale. *Pain* 2001; 94(2): 149-158.

[119] de Vet, HC, Foumani, M, Scholten, MA *et al.* Minimally important change values of a measurement instrument depend more on baseline values than on the type of intervention. *Journal of Clinical Epidemiology* 2015; 68(5): 518-524.

[120] Nicolini, A, Barbagelata, E, Tagliabue, E *et al.* Gender differences in chronic obstructive pulmonary diseases: a narrative review. *Panminerva Medicine* 2018; 60(4): 192-199.

[121] Dankers, M, Nelissen-Vrancken, MHJMG, Surminski, SMK *et al*. Healthcare Professionals' Preferred Efficacy Endpoints and Minimal Clinically Important Differences in the Assessment of New Medicines for Chronic Obstructive Pulmonary Disease. *Frontiers in Pharmacology* 2020; 10: 1519.

Chapter 9

Abbreviations

Δ	Change score
0.5SD	Half standard deviation
6MWD	Six minute walking distance
95%CI	95% confidence interval
A	Activity score on the SGRQ
ACOS	Asthma COPD overlap syndrome
AECOPD	Acute exacerbation chronic obstructive pulmonary disease
AQ	Airway Questionnaire
ATS	American Thoracic Society
AUC	Area under the curve
BDI	Baseline Dyspnoea Index
BCSS	Breathlessness Cough and Sputum Scale
BLVR	Bronchoscopic lung volume reduction
BMI	Body mass index
BPQ	Breathing Problems Questionnaire
BTS	British Thoracic Society
CAL	Chronic airflow limitation
CAT	COPD Assessment Test
CCQ	Clinical COPD Questionnaire
CDLM	Capacity of Daily Living during the Morning Questionnaire
CI	Confidence interval
CID	Clinically important difference
COPD	Chronic obstructive pulmonary disease
COPD-SIB	COPD Specific Item Bank
CRQ	Chronic Respiratory Questionnaire
CSD	Cough Severity Diary
D	Dyspnoea score on the CRQ
DALY	Disability adjusted life year
DartmCoop	Dartmouth Northern New England Primary Care Cooperative Information Project chart system
DMQ-CAT	Dyspnoea Questionnaire Computer Adaptive Test
E	Emotion score on the CRQ
ED	Emergency department
EQ-5D	EuroQol-5D
EQ-5D-3L-UI	EuroQol 5 Dimensions 3 Levels Utility Index
EQ-5D-3L-VAS	EuroQol 5 Dimensions 3 Levels Visual Analogue Scale
EQ-5D-5L-UI	EuroQol 5 Dimensions 5 Levels Utility Index
EQ-5D-5L-VAS	EuroQol 5 Dimensions 5 Levels Visual Analogue Scale
ERS	European Respiratory Society
ES	Effect size
EU	European Union
F	Functional score on the CCQ or fatigue score on the CRQ
FDR	False discovery rate
FEV_1	Forced expiratory volume in one second
FEV_1%pred	Forced expiratory volume in one second percentage predicted
FT	Feeling Thermometer
FVC	Forced vital capacity
GIC	Global impression of change
GPE	Global perceived effect

GOLD	Global initiative for obstructive lung disease
GRC	Global rating of change
GRIAC	Groningen research institute for asthma and COPD
GRIAC-PC	Groningen research Institute for asthma and COPD primary care group
GCSQ	Global Chest Symptoms Questionnaire
HRQoL	Health-related quality of life
HS-COPD	Health States COPD
I	Impact score on the SGRQ
ICC	Intraclass correlation coefficient
IMT	Inspiratory muscle training
IQR	Inter quartile range
JSM	Junior Scientific Masterclass
LAS/VAS	Linear Analogue Scale/Visual Analogue Scale
LCOPD	Living with COPD Questionnaire
LCADL	London Chest Activity of Daily Living Questionnaire
LCQ	Leicester Cough Questionnaire
LTOT	Long-term oxygen therapy
M	Mental score on the CCQ or mastery score on the CRQ
MCD	Minimal clinical difference
MCID	Minimal clinically important difference
MCSD	Minimal clinically significant difference
MDC	Minimum detectable change
MDC95	Minimal detectable change 95%
MDD	Minimally detectable difference
MIC	Minimally important change
MID	Minimum important difference
mMRC	Modified Medical Research Council Dyspnoea Scale
MPD	Minimally perceptible difference
MRF-28	Maugeri Respiratory Failure 28 items
MYMOP	Measure Yourself Medical Outcome Profile
N	Number of patients
NA	Not applicable
NHG	Nederlands Huisartsen Genootschap
NHP	Nottingham Health Profile
NR	Not reported
NS	Not significant
OSAS	Obstructive sleep apnoea syndrome
PFSDQ	Pulmonary Functional Status and Dyspnoea Questionnaire
PFSS	Pulmonary Functional Status Scale
PR	Pulmonary rehabilitation
PRISMA	Preferred reporting items for systematic reviews and meta-analyses
PRO	Patient-reported outcomes
QoL	Quality of life
QOLRIQ	Quality of Life for Respiratory Illness Questionnaire
QWBSA	Quality of Well Being Self-Administered
R	Range
R^2	Proportion of the variance of the dependent factor explained by independent factors
RCI	Reliability of change index
RCP	Routine clinical practice

RCT	Randomised controlled clinical trial
ROC	Receiver operating characteristics
RIMTCORE	Routine inspiratory muscle training within COPD rehabilitation
RQLQ	Respiratory Quality of Life Questionnaire
RV	Residual volume
S	Symptoms score on the CCQ or SGRQ
SBOH	Employer of general practitioner trainees
SD	Standard deviation
SDD	Smallest detectable difference
SEM	Standard error of measurement
Sens	Sensitivity
SF-6D	Short-Form 6 Dimensions
SF-12	Short-Form-12
SF-36	Short-Form-36
SF-CRQ	Short-Form Chronic Respiratory Questionnaire
SGRQ	St. George's Respiratory Questionnaire
SIP	Sickness Impact Profile
SMD	Standardised mean difference
SOBDA	Shortness of Breath with Daily Activities Questionnaire
Spec	Specificity
SRI	Severe Respiratory Insufficiency Questionnaire
SRM	Standardised response mean
SSD	Subjectively significant difference
SWT	Shuttle walking test
T	Total score on the CCQ, CRQ or SGRQ
T0	Time-point 0: Baseline measurement
TI	Time-point 1: Post-rehabilitation at 3 weeks
T2	Time-point 2: 3-months follow-up
T3	Time-point 3: 6-months follow-up
T4	Time-point 4: 9-months follow-up
T5	Time-point 5: 12-months follow-up
TDI	Transition Dyspnoea Index
TR	Transition rating
UCSD-SOBQ	University of California San Diego Shortness of Breath Questionnaire
UI	Utility Index
UK	United Kingdom
UMCG	University Medical Center Groningen
USA	United States of America
VAS	Visual Analogue Scale
VSRQ	Visual Simplified Respiratory Questionnaire
WHOQOLBREF	World Health Organisation Quality of Life short version list
WPAI-COPD	Work Productivity and Activity Impairment Questionnaire
YLD	Years lived with disability
Yrs	Years

Chapter 10

Nederlandse samenvatting

10.1 Publieke samenvatting

10.1.1 Achtergrond en doelstellingen

De arts en wetenschapper kunnen diverse medische onderzoeken uitvoeren, zoals het meten van de bloeddruk, bloedprikken en een blaastest voor de longen. Deze lichamelijke testen zeggen echter weinig over de door de patiënt ervaren kwaliteit van leven. Deze kwaliteit van leven wordt vaak niet bepaald door lichamelijke afwijkingen, maar door de combinatie van ervaren symptomen, invloed op het leven en beperkingen door een ziekte. Dit wordt ook wel iemands *gezondheidsstatus* genoemd. Er bestaan heel veel vragenlijsten om gezondheidsstatus te meten. Voor elke test of vragenlijst geldt in ieder geval dat er bepaald moet kunnen worden: a) wanneer een patiënt zich daadwerkelijk beter of slechter voelt na een behandeling, en b) welke verandering in de score daarbij hoort. Soms is de gemeten verandering namelijk onvoldoende belangrijk voor de patiënt. Dit zijn dan de normale schommelingen in de testuitslag. Net zoals dat iemand de ene dag meer kan wegen dan de andere, zonder dat iemand echt is afgevallen of aangekomen. De veranderscore, die wel past bij een echte ervaren verbetering of verslechtering, noemen we de drempelwaarde voor klinisch relevante verandering. In het Engels wordt dit de *minimal clinically important difference* genoemd. De afkorting hiervoor is *MCID*.

In dit proefschrift hebben de onderzoekers de MCID's van meerdere vragenlijsten opnieuw berekend, met elkaar vergeleken en gekeken welke factoren daarop van invloed waren. Het onderzoek is specifiek uitgevoerd bij mensen met de longziekte chronic obstructive pulmonary disease (COPD) met de drie belangrijkste vragenlijsten om gezondheidsstatus bij die ziekte te meten. De namen van deze drie vragenlijsten zijn de *COPD Assessment Test (CAT), Clinical COPD Questionnaire (CCQ)* en *St. George's Respiratory Questionnaire (SGRQ)*. Bij mensen met de longziekte COPD zien we dat deze vragenlijsten veel worden gebruikt door de arts, en ook verplicht zijn binnen het wetenschappelijke onderzoek.

10.1.2 Belangrijkste bevindingen

In dit proefschrift werden gegevens verzameld vanuit drie verschillende onderzoeken: (1) een analyse van de studies, die door anderen al gedaan zijn (*een literatuuronderzoek*); (2) een studie bij mensen met COPD in een Duitse longrevalidatiekliniek; en (3) een vragenlijstonderzoek bij mensen met COPD onder behandeling bij de huisarts en/of longarts in Nederland.

Allereerst, uit het literatuuronderzoek kwam naar voren dat er voor 12 verschillende vragenlijsten voor mensen met COPD al MCID's zijn bepaald. De kwaliteit van deze onderzochte studies was wisselend *(Hoofdstuk 3)*. Vooral de MCID van de vragenlijst

St. George's Respiratory Questionnaire was niet goed bepaald. Grofweg is de drempelwaarde voor klinisch relevante verandering *(een MCID)* 7 tot 10% op de totaalscore van een vragenlijst. Dit proefschrift laat dan ook zien dat de in de literatuur algemeen erkende MCID van de meest gebruikte vragenlijst - de SGRQ - 7 punten is op een schaal van maximaal 100 punten. Dit in plaats van de algemeen aangenomen waarde van 4 punten. De gemeten MCID's van de CCQ (0.40-0.50) en CAT (2-3) komen min of meer overeen met de in de literatuur gevonden waarden. Aangezien de effecten van veel medicijnen en behandelingen voor mensen met de longziekte COPD voornamelijk zijn gemeten met de SGRQ en de hierbij aangenomen drempelwaarde (MCID) van 4 punten in plaats van 7 punten, wordt de invloed van deze behandelingen op de kwaliteit van leven van de patiënt systematisch overschat. Met andere woorden, geneesmiddelen en behandelingen lijken effectief op basis van die vragenlijst, maar zijn dit wellicht toch niet.

Ten tweede, in de Duitse longrevalidatiestudie en in het Nederlandse vragenlijstonderzoek, zagen de onderzoekers dat er een heleboel verschillende methodieken gebruikt kunnen worden om een MCID te berekenen. Er werden dan ook kleine verschillen gevonden tussen de resultaten van de gebruikte methodieken *(Hoofdstuk 4)*. De onderzoekers zagen ook enkele kleine verschillen in de vragenlijst MCID's, wanneer ze deze berekenden gedurende verschillende periodes van verandering (*Hoofdstuk 5)*. De meetperiodes varieerden daarbij van drie weken tot een jaar. Ondanks de kleine verschillen, bleken de vragenlijst MCID's voor mensen met een verbetering in de gezondheidsstatus, redelijk goed vergelijkbaar met de MCID's voor mensen met een ervaren verslechtering in de gezondheidsstatus (*Hoofdstuk 6)*.

Ten derde, de onderzoekers beoordeelden verschillende factoren, die van invloed waren op de waarde van een MCID. Er werd ontdekt dat MCID's voor mensen met de longziekte COPD in de Duitse longrevalidatiestudie veel hoger waren dan voor de Nederlandse patiënten bij de huisarts en de longarts (*Hoofdstuk 7)*. Daarnaast werd een grotere verbetering in gezondheidsstatus (*een hogere MCID*) gezien bij deelnemers, die al veel klachten en beperkingen hadden aan het begin van het onderzoek, en die zich daadwerkelijk beter voelden na beide studies. Dat was ook zo voor vrouwen, deelnemers ouder dan 60 jaar, patiënten met een goede longfunctie, en mensen met minder andere bijkomende ziektes. Echter, deze deelnemers hadden juist weinig aanvullende verslechtering in gezondheidsstatus nodig (*een lagere MCID*) om zich daadwerkelijk slechter te voelen. De MCID's van de drie COPD vragenlijsten in dit proefschrift konden niet volledig verklaard worden door een complex model met al deze verschillende factoren.

10.1.3 Conclusie

Het is goed om te beseffen dat MCID's voor gezondheidsstatusvragenlijsten een verplichte ondergrens zijn voor de goedkeuring en evaluatie van nieuwe medicijnen en andere behandelingen. Het is daarom dan ook erg belangrijk dat er goede MCID's voor testen en vragenlijsten worden berekend. Het zou ook waardevol kunnen zijn om verschillende MCID's in de praktijk te gebruiken voor mensen met de longziekte COPD op basis van hun eigenschappen en de onderzoekssituatie. Op die manier kan er een meer individueel oordeel worden gegeven aan de gemeten verandering. In de discussie van dit proefschrift presenteren de onderzoekers een dynamisch model voor het gebruik van de MCID's van de drie belangrijkste COPD vragenlijsten. Ook worden aanbevelingen gedaan hoe andere onderzoekers en artsen goede MCID's kunnen bepalen voor hun (nieuwe) vragenlijsten en medische testen.

10.2 Academische samenvatting

10.2.1 Achtergrond en doelstellingen proefschrift

De primaire doelstelling van dit proefschrift was om meer inzicht te verkrijgen in de dynamiek van de belangrijke internationale parameter *minimal clinically important difference* (MCID) voor vragenlijsten op het gebied van gezondheid-gerelateerde kwaliteit van leven (ook wel *gezondheidsstatus*) van patiënten met de chronische longziekte *chronic obstructive pulmonary disease* (COPD). Het bepalen van gezondheidsstatus behelst een gestandaardiseerde manier om de impact van ziekte en gezondheid te meten op het welzijn van de patiënt in het dagelijks leven. Uit eerder wetenschappelijk onderzoek is gebleken dat bekende fysiologische uitkomstmaten, zoals bloedonderzoek en diagnostische testen, slechts beperkt correleren met de door de patiënt ervaren symptomen, functionele mogelijkheden en kwaliteit van leven. Dit is met name actueel voor patiënten met een (progressieve) chronische ziekte, zoals COPD. Gedurende de afgelopen decennia zijn er dan ook diverse instrumenten ontwikkeld om gezondheidsstatus te meten (*Hoofdstuk 1 en 2)*.

COPD is wereldwijd één van de meest frequent voorkomende aandoeningen, gekenmerkt door een diversiteit aan luchtwegklachten zoals dyspneu, hoesten en overmatige slijmvorming. Deze symptomen worden veroorzaakt door een chronische ontsteking van de luchtwegen, resulterend in onomkeerbare schade. Voor patiënten met de longziekte COPD zijn de belangrijkste gevalideerde gezondheidsstatusvragenlijsten de *COPD Assessment Test* (*CAT*, 8 vragen, 1 domein, score 0-40 punten), de *Clinical COPD Questionnaire* (*CCQ*, 10 vragen, 3 domeinen, score 0-6 punten) en *de St. George's Respiratory Questionnaire* (*SGRQ*, 50 vragen, 3 domeinen, score 0-100 punten). Voor deze drie vragenlijsten geldt: hoe hoger de score, hoe slechter de gezondheidsstatus van de

patiënt (*Hoofdstuk 2)*. Een negatieve veranderscore representeert daarmee dan ook een verbetering.

Gezondheidsstatusvragenlijsten zijn tegenwoordig een verplichte uitkomstmaat in wetenschappelijke studies en worden frequent toegepast in de klinische praktijk. Belangrijk voor welke uitkomstmaat dan ook is dat deze geïnterpreteerd moet kunnen worden door de arts en onderzoeker. Hierbij geldt dat er ten aanzien van de gemeten verandering na een behandeling onderscheid gemaakt moet worden tussen toevallige verandering (*ruis)* en daadwerkelijk relevante verandering (*signaal)*. Een MCID kan hier tussen differentiëren. De parameter is in de jaren `90 geïntroduceerd als de minimale drempelwaarde waarbij de gemeten verandering na therapie als klinische relevant kan worden beschouwd voor de patiënt. Dit rechtvaardigt het succes van de gekozen behandeling. Sinds de ontwikkeling van het concept, wordt de MCID van een instrument frequent toegepast in wetenschappelijke studies om de gemeten veranderingen te evalueren en te interpreteren (*Hoofdstuk 1*).

Tot op heden is er nog weinig duidelijkheid omtrent de diverse dynamische factoren, die mogelijk van invloed zijn op de hoogte van een MCID. Zo zijn er onduidelijkheden omtrent welke methodologie (*anker-gebaseerde, statistische of opinie-gebaseerde technieken*) gebruikt dient te worden voor het bepalen van de MCID van een instrument. Daarnaast is het onduidelijk in hoeverre de duur van de periode voor het meten van de verandering (de zogeheten *follow-up periode*) van invloed is op de MCID. Factoren zoals *recall bias* (moeite met het herinneren van de vorige gezondheidsstatus ter vergelijking) en *response shift* (een verandering in de interpretatie van de gezondheidsstatus) zouden hierbij van invloed kunnen zijn. Ten derde is het onduidelijk of MCID's voor verbetering vergelijkbaar zijn met die voor de interpretatie van verslechtering in gezondheidsstatus. Tot slot, is er maar weinig bekend in hoeverre context- en patiënt-gerelateerde factoren van invloed zouden kunnen zijn op de MCID van een instrument (*Hoofdstuk 1*). Dit proefschrift geeft meer inzicht in deze aspecten.

In dit proefschrift worden in de analyses data gebruikt uit drie studies. Het betreft (1) een systematisch literatuuronderzoek en meta-analyse omtrent de MCID van gezondheidsstatusvragenlijsten voor patiënten met COPD; (2) een gerandomiseerde klinische studie naar de effectiviteit van additionele ademhalingstherapie (zogeheten *inspiratory muscle training* (IMT)) tijdens drie weken longrevalidatie van 451 COPD patiënten in Duitsland; en (3) een observationele studie met 207 Nederlandse COPD patiënten in de reguliere eerste- en tweedelijns zorg gedurende 12 maanden zonder additionele interventie (*Hoofdstuk 1*).

10.2.2 Belangrijkste resultaten

De huidige gebruikte MCID's zijn voor de CAT 2 punten, voor de CCQ 0.40 punten en voor de SGRQ 4 punten. In het systematische literatuuronderzoek in de databases PubMed, EMBASE en de Cochrane Library werden vanuit 785 artikelen uiteindelijk 21 publicaties geïdentificeerd over de MCID van 12 verschillende gezondheidsstatusvragenlijsten voor patiënten met COPD (*Hoofstuk 3*). Op basis van de meta-analyse waren de gewogen MCID's voor verbetering als volgt: -2.54 voor de CAT (6 publicaties, bereik -3.80 tot -1.00), -0.43 voor de CCQ (5 publicaties, bereik -0.62 tot -0.21), en -7.43 voor de SGRQ (4 publicaties, bereik -10.19 tot -2.40) (*Hoofdstuk 8, Tabel 1, Figuren 1-3)*. Voor de andere 9 instrumenten waren de geïncludeerde publicaties te heterogeen of te weinig in aantal. Er werden geen duidelijke patronen waargenomen voor een eventuele invloed van studie, context en duur van de follow-up periode van de geïncludeerde studies op de hoogte van de MCID. Het bewijs voor de MCID's van de CAT en CCQ bleek sterk en de gewogen schattingen werden beoordeeld als valide. De momenteel in de praktijk toegepaste MCID voor de SGRQ (4 punten), maar ook die van een andere vragenlijst, de Chronic Respiratory Questionnaire (CRQ, 0.50 punten), kwamen slecht overeen met de onderzochte inhoud. Deze MCID's bleken in werkelijkheid veel hoger te zijn. De toepassing van een te lage MCID als drempelwaarde, kan leiden tot een overschatting van de interpretatie van de effecten van behandelingen voor patiënten met COPD in wetenschappelijke studies en de dagelijkse praktijk. Tot slot, MCID's voor verslechtering waren schaars en toonden aan dat daar meer onderzoek voor nodig is.

Op basis van de geanalyseerde data van de 451 COPD patiënten tijdens het longrevalidatietraject in Duitsland (gemiddelde leeftijd 58 jaar, 65% man) konden diverse methoden vergeleken worden voor de bepaling van de MCID van de CAT, CCQ en SGRQ (*Hoofstuk 4)*. Ook werden de MCID's voor de domeinscores van de CCQ en SGRQ voor het eerst beoordeeld. De bepaalde MCID's verschilden afhankelijk van de gebruikte technieken en dit resulteerde in een diversiteit van schattingen (*Hoofdstuk 8, Tabel 1, Figuren 1-3)*. De gewogen MCID's waren -3.28 voor de CAT (bereik -6.43 tot -1.46), -0.52 voor de CCQ (bereik -0.62 tot -0.28), en -7.91 voor de SGRQ (bereik -10.19 tot -5.20). Over het algemeen waren de resultaten vanuit de verschillende anker-gebaseerde technieken redelijk vergelijkbaar. De ankers ter referentie betroffen hier (1) een globale schatting van de ervaren verandering voor de patiënt; (2) een COPD exacerbatie als criterium voor een klinische gebeurtenis; (3) de andere COPD gezondheidsstatusvragenlijsten met de huidige gebruikte MCID's. Een zorgvuldige selectie van ankers ter referentie moet echter wel worden overwogen, zeker wanneer diens MCID's betwist zouden kunnen worden.

De MCID op basis van de statistische techniek halve standaard deviatie (0.50SD) was het beste te vergelijken met de anker-gebaseerde technieken (*Hoofdstuk 4*). De methode

van de standard error of measurement (SEM) was inconsistent, en de 1.96SEM was zeer conservatief voor de CAT en de SGRQ. De MCID's voor de domeinscores op de CCQ en SGRQ waren redelijk vergelijkbaar met de waarden voor de totale vragenlijstscore, met uitzondering van de mentale score op de CCQ en de symptoomscore op de SGRQ. De meeste MCID schattingen voor de CAT en CCQ waren vergelijkbaar of net iets hoger dan de geaccepteerde waarden vanuit de literatuur (respectievelijk 2 en 0.40 punten). Echter, de MCID's voor de SGRQ waren allemaal fors hoger dan de geaccepteerde drempel van 4 punten. Deze drempel wordt uitgebreid toegepast in de huidige wetenschappelijke literatuur. Dit zou kunnen hebben geresulteerd in een overschatting van de interpretatie van de gevonden behandeleffecten. Samenvattend, klinisch relevante verbetering dient over het algemeen te worden overwogen bij een verandering van 3 punten op de CAT, 0.50 punten op de CCQ en 7 punten op de SGRQ. Deze voorgestelde drempels zijn grofweg 7% van de maximale totaalscore.

Gegevens van de longrevalidatiepatiënten in Duitsland werden ook gebruikt om te bepalen of de MCID's van de CAT, CCQ en SGRQ verschilden wanneer deze bepaald werden direct na de interventie en tijdens verschillende follow-up periodes ten gevolge van eventuele *recall bias* (*Hoofdstuk 5*). Van de 451 patiënten, voltooiden 309 patiënten de gehele follow-up na 1 jaar. Als anker-gebaseerde techniek werden een ankervraag (*global rating of change*) gebruikt met 15 antwoordopties en één met 5 antwoordopties, waarbij de patiënt werd gevraagd om retrospectief te beoordelen hoe de huidige gezondheidsstatus was in vergelijking tot het begin van de interventie. Globaal gezien werden er geen significante verschillen gevonden tussen de MCID schattingen voor verbetering na een follow-up periode van 3 weken, en 3, 6, 9 en 12 maanden bij gebruik van de ankervraag met 15 antwoordopties. De MCID's voor verbetering varieerden van -3.1 tot -2.3 voor de CAT; -0.6 tot -0.4 voor de CCQ; en -10.3 tot -7.6 voor de SGRQ (*Hoofdstuk 8, Tabel 1, Figuren 1-3*). Hogere (niet-significante) MCID's werden waargenomen voor de CAT en CCQ met een follow-up periode van 3 weken direct na afronding van het longrevalidatietraject. De waarde van de MCID van een instrument kan mogelijk dus hoger uitvallen, wanneer deze bepaald wordt direct na een interventie met een kortere follow-up periode.

Echter, het gebruik van een ankervraag met slechts 5 antwoordopties resulteerde wel in significant lagere MCID's voor de CAT en CCQ na 12 maanden follow-up in vergelijking tot de ankervraag met 15 antwoordopties (*Hoofdstuk 5*). De MCID schattingen op basis van deze alternatieve ankervraag waren -1.4 voor de CAT (significant verschil -1.4), -0.3 voor de CCQ (significant verschil -0.2), en -7.7 voor de SGRQ (niet-significant verschil -1.1) (*Hoofdstuk 8, Tabel 1, Figuren 1-3)*. De classificatie van de geïncludeerde patiënten volgens beide ankervragen kwam voor 55% overeen. Kennelijk resulteerde het hebben van minder

antwoordopties op een ankervraag in een lagere MCID. Minder antwoordopties zou kunnen leiden tot verlies van belangrijke informatie en een verminderd discriminerend vermogen van de schaal.

In *Hoofdstuk 6* werden data van de beide klinische studies in Duitsland en Nederland gebruikt om te beoordelen in hoeverre MCID's voor verbetering in gezondheidsstatusvragenlijsten vergelijkbaar waren met die voor verslechtering. Het voorkomen van verslechtering bij een progressieve en chronische aandoening zoals COPD, kan een belangrijk resultaat van behandeling zijn. Hiertoe is het noodzakelijk klinisch relevante verslechtering van willekeurige variaties (*ruis*) te onderscheiden. Het systematische literatuuronderzoek (*Hoofdstuk 3)* toonde aan dat MCID's voor verslechtering niet bekend zijn voor de CAT, CCQ en SGRQ. Op dit moment worden diens MCID's voor verbetering simpelweg ook toegepast voor de interpretatie van verslechtering. Bij de beoordeling van de data van 451 COPD patiënten tijdens hun longrevalidatie en 207 COPD patiënten in de reguliere medische zorg, bleek dat de anker-gebaseerde en statistisch-gebaseerde MCID's voor verbetering en verslechtering redelijk overeen kwamen. De MCID schattingen verschilden wel tussen de Duitse longrevalidatiegroep en de patiënten zonder aanvullende interventie in de reguliere Nederlandse medische zorg.

Op basis van de huidige data gedurende diverse follow-up momenten, werden de volgende drempelwaarden gevonden voor minimale relevante verbetering dan wel verslechtering: CAT≥3 (interventie) en CAT≥2 (reguliere medische zorg); CCQ ≥0,40 (interventie) en CCQ≥0,30 (reguliere medische zorg), SGRQ≥6 (interventie) en SGRQ ≥5 (reguliere medische zorg) (*Hoofdstuk 6, Hoofdstuk 8, Tabel 1, Figuren 1-3)*. De gevonden MCID schattingen voor de CAT en CCQ kwamen goed overeen met eerdere resultaten vanuit de literatuur. De MCID schattingen voor de SGRQ vanuit de longrevalidatiegroep waren wederom fors hoger dan de voorgestelde en toegepaste drempel van 4 punten in de literatuur. De MCID schattingen voor COPD patiënten gedurende de reguliere medische zorg waren deels vergelijkbaar met deze drempelwaarde van 4 punten op de SGRQ. Daarnaast zouden mogelijk de afkappunten voor middelmatige en grote klinisch relevante verandering kunnen liggen in het bereik van respectievelijk 4-5 en 5-6 punten op de CAT; 0.80 en 1.00 punten op de CCQ; en 10-15 en 15-20 punten op de SGRQ. Meer onderzoek hiernaar is echter noodzakelijk, gezien de kleine hoeveelheid patiënten met een dergelijke grote verandering.

Tot slot werd in dit proefschrift onderzocht in hoeverre patiënt-gerelateerde factoren, onderzoekssituatie en de ernst van het beginniveau van de gezondheidsstatus van invloed waren op de hoogte van de MCID van de drie gezondheidsstatusvragenlijsten voor patiënten met COPD (*Hoofdstuk 7)*. Data van 658 COPD patiënten tijdens

longrevalidatie of reguliere medische zorg werden retrospectief geanalyseerd (2299 veranderscores) met behulp van anker-gebaseerde en statistische methodieken. De meeste MCID schattingen voor verbetering en verslechtering lagen voor de CAT tussen de ±1.50 en ±3.50; voor de CCQ tussen de ±0.30 en ±0.60; en voor de SGRQ tussen de 4 en 9 punten (*Hoofdstuk 8, Tabel 1, Figuren 1-3)*. Vanuit de subgroepanalyses werden diverse trends gezien. De drempelwaardes voor klinisch relevante verandering waren significant hoger voor de interventiegroep in vergelijking tot de reguliere medische zorg patiënten. Daarnaast waren de MCID's voor verbetering op de CAT, CCQ en SGRQ significant drie- tot zevenmaal hoger voor patiënten met vanaf het begin een slechtere gezondheidsstatus; echter de MCID's voor verslechtering waren vier- tot zesmaal kleiner in vergelijking tot de groep patiënten met een betere gezondheidsstatus. Vrouwen, patiënten ouder dan 60 jaar, patiënten met een betere longfunctie (GOLD categorie I-II), en patiënten met minder comorbiditeiten hadden (niet-significante) hogere MCID schattingen voor verbetering. Zij hadden daarentegen kleinere MCID's voor verslechtering in vergelijking tot hun counter-groep.

Multipele lineaire regressiemodellen toonden dat de ernst van het beginniveau van de gezondheidsstatus en de onderzoekssituatie belangrijke significante onafhankelijke factoren waren voor de hoogte van de MCID. Er werden echter complexe interacties waargenomen tussen de diverse variabelen en mogelijk ook invloed van het regressie naar het gemiddelde fenomeen. De verklaarde variantie van de modellen was laag. Hoewel de MCID op dit moment gezien wordt als een statische parameter voor de interpretatie van behandeleffecten, speelt een complexe interactie van setting, niveau van de gezondheidsstatus en patiëntfactoren een rol in de uitkomstwaarde. Indien een individuele interpretatie van verandering wenselijk is in het wetenschappelijke onderzoek of in de klinische praktijk, zou men baat kunnen hebben bij geclusterde of zelfs gepersonaliseerde MCID's gebaseerd op de diversiteit aan dynamische factoren. Hiertoe is nader onderzoek wenselijk met grote aantallen patiënten uit verschillende populaties.

10.2.3 Discussie en conclusie

Op basis van bovenstaande resultaten werden alle MCID's geschat voor de CAT tussen de -6.43 en -0.67 voor *verbetering,* en tussen de 0.50 en 6.30 voor *verslechtering.* Voor de CCQ waren deze schattingen -0.82 tot -0.10 voor *verbetering,* en 0.19 tot 0.84 voor *verslechtering*. En voor de SGRQ waren deze waardes -12.28 tot -2.40 voor *verbetering*, en 0.33 tot 12.86 voor *verslechtering (Hoofdstuk 8, Tabel 1, Figuur 1-3).* Echter, de meeste van deze schattingen voor verbetering en verslechtering waren tussen de ±2.00 en ±3.50 voor de CAT, en tussen de ±0.30 en ±0.50 voor de CCQ. Voor de SGRQ waren de meeste MCID's voor verbetering tussen de -9.00 en -6.00, en voor verslechtering tussen de +5.00 en +8.00.

De MCID is en blijft een complexe parameter, die cruciaal is voor de interpretatie van behandeleffecten. Hoewel in dit proefschrift de dynamiek niet volledig is opgehelderd, zijn er weinig alternatieven voor het evalueren en interpreteren van verandering. Als zodanig blijft de MCID van vitaal belang, maar moet ook met voorzichtigheid worden behandeld. Op dit moment wordt de MCID als statische vaste drempelwaarde gebruikt zonder enige flexibiliteit. Daarom wordt in dit proefschrift een model gepresenteerd op basis van de diverse schattingen voor de dynamische MCID evaluatie van gezondheidsstatus in patiënten met COPD. Hierbij wordt niet alleen gekeken naar de resultaten van een onderzoek en het betrouwbaarheidsinterval, maar kan de interpretatie ook verschuiven op basis van de onderzoekssituatie en steekproefkenmerken *(Hoofdstuk 8, Tabel 2, Figuur 4)*.

De bepaling van de MCID van een instrument vereist dat een structurele en uniforme aanpak wordt gevolgd en dient niet als vanzelfsprekend te worden beschouwd. Daarom worden in dit proefschrift meerdere aanbevelingen gedaan voor de onderzoeker en arts om op een gestructureerde manier een MCID te bepalen voor een instrument (*Hoofdstuk 8, Box 2)*. Hierbij worden drie kwaliteitsniveaus gepresenteerd: brons, zilver en goud. De aanbevelingen gelden niet alleen voor de MCID's van vragenlijsten, maar eigenlijk voor elke diagnostische test waarmee een verandering kan worden gemeten. Dit kan zowel betrekking hebben op nieuwe als bestaande instrumenten. Wanneer een kwalitatieve aanpak niet gevolgd wordt, kan dit leiden tot foutieve MCID's waarbij er een over- of onderschatting volgt van de interpretatie van de uitkomsten van wetenschappelijke studies en behandeling in de kliniek.

Tot slot moet er kritisch gekeken worden naar de interpretatie van bestaande studieresultaten voor de behandeling van COPD. Dit is noodzakelijk, aangezien de MCID van met name de SGRQ hoger zou moeten zijn dan de huidige gebruikte waarde van 4 punten. Artsen, onderzoekers en instituten voor de beoordeling van geneesmiddelen moeten kritisch blijven en zich bewust zijn van de achtergrond van de gebruikte MCID's voor de interpretatie van behandeleffecten.

Chapter 11

Dankwoord

In 2011 maakte ik voor het eerst kennis met het uitvoeren van wetenschappelijk onderzoek binnen de afdeling huisartsgeneeskunde van het UMCG als onderdeel van de Junior Scientific Masterclass (JSM) van de Rijksuniversiteit Groningen. Als beginnend student geneeskunde mocht ik helpen data te verzamelen binnen het IMI ProActive project. Het contact met jullie, **Corina en Thys**, resulteerde in een stage wetenschap waarin ik de basis kon leggen voor mijn promotietraject. Een klein jaar later bood de JSM mij de mogelijkheid om mijn onderzoek en het promotietraject daadwerkelijk te starten. Ik wil dan ook de **JSM** hartelijk danken voor de financiering van mijn MD-PhD traject naar de minimal clinically important difference (MCID) van gezondheidsstatusvragenlijsten voor patiënten met COPD.

Uiteraard wil ik allereerst mijn promotieteam enorm bedanken voor de samenwerking, overleggen, feedback, plezier en kennisoverdracht. **Thys**, ik weet nog goed dat je destijds opperde om iets te gaan doen met de MCID van vragenlijsten bij COPD. Jij stelde voor om dat vooral eens tot op de bodem uit te zoeken, want dat was nog niet goed gedaan. Op dat moment had ik nog geen flauw idee wat een MCID was, wat voor belang het zou dienen en waarom dokters eigenlijk vragenlijsten gebruiken bij patiënten met COPD. Inmiddels ben ik overtuigd van de toegevoegde waarde ervan. Heel veel dank Thys, als eerste promotor, voor je initiële aanzet tot dit onderzoek. En ook voor jouw kennis en netwerk, waardoor het proces uiteindelijk heeft geresulteerd in dit proefschrift. Samen hebben we een brede invulling kunnen geven aan het onderwerp met mooie resultaten. Jij met name kijkend als huisarts, hoogleraar en pragmaticus; ik misschien iets meer theoretisch. Ik heb de samenwerking met jou als zeer prettig, informeel en open ervaren. Ik ben trots dat ik als promovendus bij jou heb mogen groeien. Ik wens je veel gezondheid en plezier tijdens je welverdiende pensioen samen met Phil, je kinderen en kleinkinderen.

Robbert, als tweede promotor ben je iets later bij het onderzoek betrokken geraakt. Binnen het team behield jij het overzicht van waaruit jij suggesties gaf hoe we iets beter konden organiseren, of meer gestructureerd konden opschrijven. Je hield me daarin scherp waardoor ik helder kreeg wat de boodschap voor de lezer moest zijn. Ik ben ervan overtuigd dat jouw rol de inhoud van dit proefschrift zeer heeft versterkt. De samenwerking met jou was prettig en open, met een dosis humor her en der. Dank voor jouw leerzame en wijze visie. Ook voor jou veel gezondheid en plezier toegewenst tijdens je pensioen samen met je familie.

Corina, als copromotor was je mijn dagelijkse aanspreekpunt en rots in de branding. Bedankt voor jouw frequente en heldere feedback, alsmede jouw begeleiding in alle facetten van het onderzoek. Je hebt een praktische instelling, tekstueel welgemeende input en uitgebreide methodologische kennis. Naast het werk aan dit proefschrift,

hebben we ook veel gezellige momenten gehad tijdens de diverse reizen en congressen. Onze samenwerking was vertrouwd, gezellig en zeker ook serieus. Ik ben ontzettend trots dat jij als copromotor het team compleet maakt. We gaan elkaar in de toekomst uiteraard nog vaker zien, en ik wens je alle geluk en gezondheid.

Tot slot, **Janwillem**, ook al ben je door omstandigheden geen officiële copromotor geworden, toch beschouw ik jou wel degelijk zo. Jouw bijdrage aan de inhoud van de stukken en onze overleggen waren absoluut van dien aard. Als huisarts en onderzoeker was jouw vraag vaak wat de huisarts hier nu mee moest. Zeker op het vlak van de MCID is het soms lastig om de praktijk dichtbij te houden. Ook heb je me gestimuleerd om de enorme hoeveelheid aan data meer visueel te presenteren. Ik heb bewondering voor je enorme diversiteit aan taken en het drukke gezinsleven. Bedankt voor onze samenwerking en we komen elkaar in het werkveld zeker nog tegen.

Naast het promotieteam in Nederland, zou het onderzoek beschreven in dit proefschrift, nooit mogelijk zijn geweest zonder de samenwerking met de medewerkers van de **Klinik Bad Reichenhall,** Center for Rehabilitation, Pulmonology and Orthopedics in Duitsland. In een gerandomiseerd onderzoek naar de werking van additionele ademhalingsoefeningen tijdens de longrevalidatie werden in grote getalen vragenlijsten afgenomen. Thys, jij legde de basis voor de samenwerking om zodoende ook verder onderzoek te kunnen doen naar de MCID van deze vragenlijsten.

First, **Konrad Schultz**, as medical director and physician, you involved me and the team in your enthusiasm for executing research (“*your little hobby*”). Thank you so much for allowing me to investigate the dynamics of the MCID within your data from pulmonary rehabilitation. It has been an honor working with you. I have great respect for your extensive energy to execute clinical research in the rehabilitation center. Thank you for your hospitality in Bad Reichenhall. I am convinced that we have collectively composed wonderful results. I hope that we can have some more time in performing additional analyses.

A warm thank you, to **all (former) colleagues within the RIMTCORE team**, especially **Michael Wittmann and Danijel Jelusic**. Thank you for educating me more on rehabilitation principles and pulmonology during my visits in Bad Reichenhall. I would also like acknowledge all (former) study nurses and physiotherapists within the RIMTCORE team. Furthermore, **Michael Schuler**, thank you for your great methodological input in the manuscripts. It has been wonderful meeting you in Würzburg. Thank you all for your input and co-authoring of most of the published manuscripts concerning the MCID.

Most certainly the data collection would not have been possible without all the **RIMTCORE patients** participating in the trial. Thank you very much for your efforts in filling out the questionnaires. I hereby would like to acknowledge specifically one of the female RIMTCORE participants, who allowed me to interview her about her personal experiences with COPD, pulmonary rehabilitation and questionnaires.

Naast de studie binnen de longrevalidatie in Bad Reichenhall, wilden we als team ook graag data vanuit de eerste en tweede lijn in Nederland toevoegen. Mijn enorme dank gaat wat dat betreft allereerst uit naar **de verschillende deelnemende huisartsenpraktijken**, waar ik samen met de huisartsen en praktijkondersteuners potentiële deelnemers heb mogen werven. Bedankt voor jullie gastvrijheid, interesse en prettige contacten. Daarbij in het bijzonder wil ik jou, **Siebrig Schokker**, bedanken voor het helpen werven van COPD patiënten. Ook wil ik het **Longfonds** hartelijk danken voor jullie hulp bij de werving van deelnemers. Daar is een zeer grote respons op gekomen vanuit heel Nederland. Tot slot, ben ik veel dank verschuldigd aan **alle deelnemers van deze Nederlandse studie**. Zonder u was het onderzoek niet mogelijk geweest. Bedankt ook voor de soms persoonlijke notities over uw ervaringen met COPD. En dank voor uw loyaliteit in het beantwoorden van alle vragen. Ik wens u alle goeds.

Graag wil ik benadrukken dat het geweldig is geweest om als MD-PhDer te worden opgenomen binnen het team van de **afdeling huisartsgeneeskunde van het UMCG.** Het is fijn om jullie als collega's te hebben ontmoet en meer te leren over de methodologische aspecten van onderzoek en het vak van huisarts. Mijn grootste dank gaat uit naar de (voormalige) **"longenclub" (GRIAC-PC)** van de afdeling. **Thys, Janwillem, Corina, Bertine, Esther, Fredrik, Job, Ellen, Roland, Hendrik-Jan, Suzanne, Anna-Jetske, Tina, Boudewijn, Ioanna, Bruce, Miguel, Onni en alle anderen**. Uiteraard ook dank voor de overdracht van kennis, maar zeker ook de gezelligheid tijdens congressen en andere bijeenkomsten. We gaan elkaar zeker zien in de toekomst! **Bertine en Ioanna**, dank ook voor jullie rol als medeauteur en feedback bij twee hoofdstukken van dit proefschrift.

Naast de GRIAC-PC groep, wil ik in het bijzonder **Boudewijn Kollen** van de afdeling huisartsgeneeskunde bedanken. Dank voor je overweldigende input op het gebied van statistiek en methodologie. Samen hebben we heel wat moeten uitzoeken en ontdekken op het gebied van de MCID-methodologie. Dank voor je enthousiasme en betrokkenheid. Dank voor al jouw wijze woorden en de mooie bijdrage aan twee van de hoofdstukken van dit proefschrift. Ik hoop dat je geniet van je welverdiende pensioen met je familie. Ook wil ik graag **Marjolein Berger, Sytse Zuidema, Gea Holtman en Huib Burger** van de afdeling huisartsgeneeskunde bedanken voor de kans, die jullie mij

hebben geboden, om verder onderzoek te kunnen blijven doen binnen de afdeling. Tot slot wil ik ook **Truus van Ittersum** (SHARE) bedanken voor jouw geweldige bijdrage aan de ontwikkeling van de zoekstrategie voor het literatuuronderzoek in dit proefschrift.

Graag wil ik ook de leden van de beoordelingscommissie - **professor Huib Kerstjens, professor Adelita Ranchor** en **professor Dave Singh** - bedanken. Thank you! Ik waardeer het enorm dat jullie tijd hebben gemaakt om mijn proefschrift te lezen en te voorzien van feedback. I appreciate it very much that you have taken the time and effort to provide my PhD thesis with interesting feedback. Ik zie uit naar het moment dat ik jullie vragen mag beantwoorden tijdens de openbare verdediging. I am looking forward to answering your questions during the public defence.

Lieve paranimfen, ik ben jullie heel erg dankbaar dat jullie mij willen bijstaan tijdens een heel belangrijk moment in mijn leven. Lieve **Stephanie**, als mede MD-PhDer hebben we veel ervaringen uitgewisseld. We hebben elkaar in het eerste blok van de studie leren kennen, en ook nu is het als we elkaar zien weer als vanouds. We kunnen alles aan elkaar vragen als arts, onderzoeker en ook op persoonlijk vlak. Ik bewonder je doorzettingsvermogen en je enthousiasme. Ik wens je alle goeds voor de toekomst samen met Arne en je familie. Lieve **Manna**, lieve zus, zowel op onderzoeksgebied bij POG als op persoonlijk vlak heb ik veel van je mogen leren. Ik kan me dan ook geen betere paranimf bedenken voor dit bijzondere moment. Bedankt voor jouw steun, liefde, gezelligheid en plezier. Ik bewonder hoe je je in alle verschillende rollen staande houdt. Je bent een hele sterke zus waar ik trots op ben. Geniet samen met je lieve gezin **Thijs, Stijn en Noortje**. Uiteraard jullie ook bedankt voor jullie liefde, vrolijkheid en interesse. Dat we in de toekomst nog veel mooie momenten samen mogen beleven.

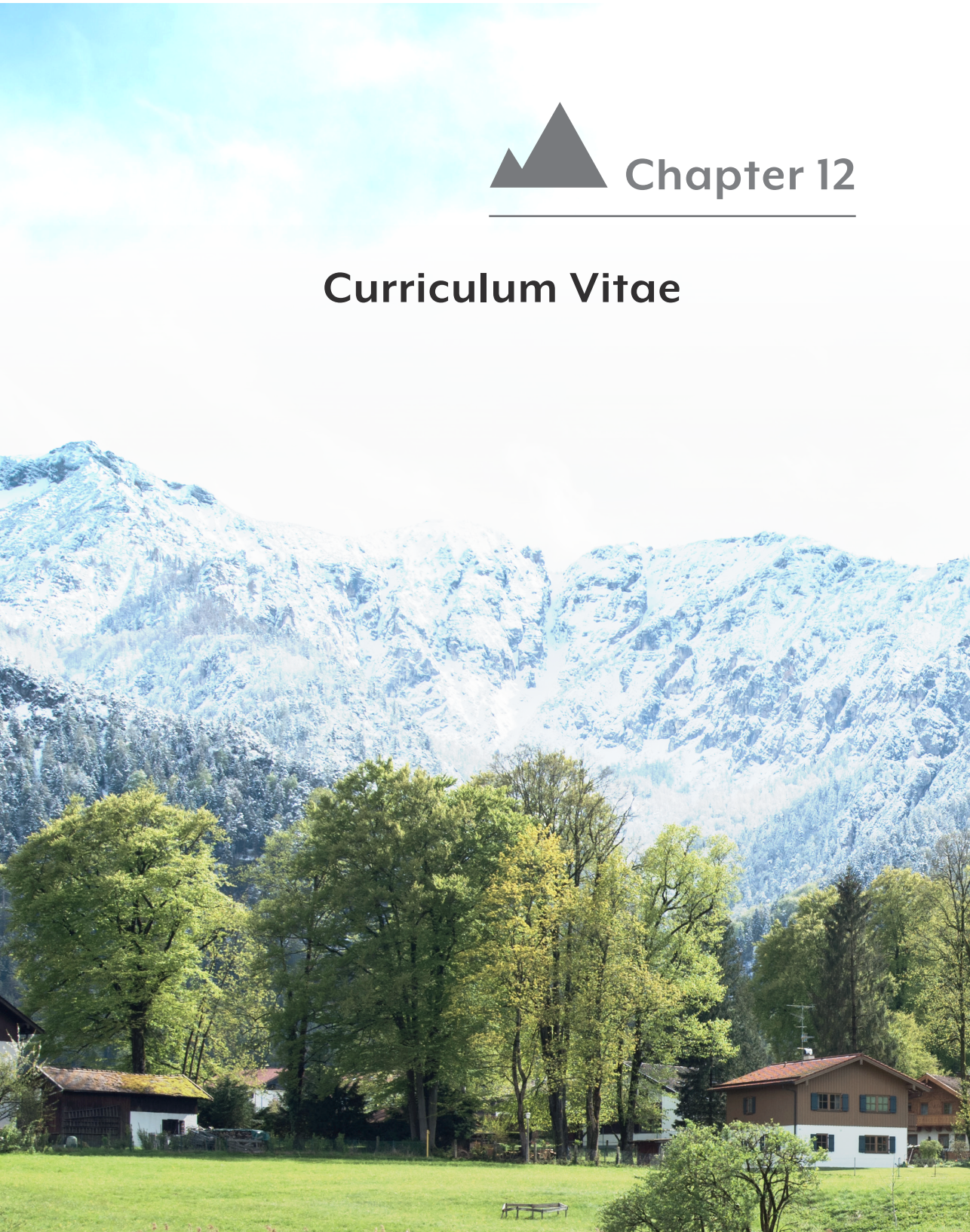
Lieve vrienden, bedankt voor jullie gezelligheid en interesse in mijn onderzoek. Het was niet altijd even makkelijk om uit te leggen waar mijn onderzoek over ging. Ik hoop dat er na de promotie weer wat meer tijd voor sociale activiteiten vrijkomt. Ik wil in het bijzonder ook **Vera** bedanken. Dank voor jouw loyale vriendschap, je lieve woorden en je oprechte steun. Ook al zien we elkaar niet vaak vanwege de afstand, het doet me goed dat ik weet dat je er bent. Alle goeds voor jou en je lieve gezin.

Lieve **pap en mam**, jullie hebben me de kans gegeven om te kunnen studeren. En ook toen ik toch nog geneeskunde wilde studeren, hebben jullie me altijd gesteund. Jullie hebben er alles aan gedaan om mijn jeugd en de tijd erna tot een geweldige tijd te maken. Het is in woorden niet uit te drukken hoe dankbaar ik ben voor jullie onvoorwaardelijke liefde en steun. Ik hoop dat jullie weten dat dit ook wederzijds is. Ik ben trots op jullie.

Lieve **Jaco**, bedankt dat je mijn mooie en oprechte broer bent. Al besef je niet waar het over gaat en ben je er niet bij tijdens de promotie, weet dat je er altijd onvoorwaardelijk bij hoort.

Lieve **schoonfamilie**, lieve Gerry, James, Rianne, Thijs, Sophie en Josephine. Bedankt voor jullie interesse, geduld, ondersteuning en gezelligheid de afgelopen jaren.

Tot slot, het allerbelangrijkste en dierbaarste in mijn leven is mijn gezin. Lieve **Daan**, toen we elkaar leerden kennen ruim tien jaar geleden, vertelde ik je dat ik nog geneeskunde wilde studeren. Je hebt mij elke seconde sindsdien ondersteund. Je hebt met mij meegedacht, geholpen pakketjes met vragenlijsten samen te stellen en vooral veel geluisterd. Zonder jou, zou dit alles niet mogelijk zijn geweest. Ik ben ervan overtuigd dat we samen elke berg kunnen beklimmen, al zitten er soms wat hobbels op de weg. Dank je wel voor jouw onvoorwaardelijke liefde, steun en geduld. Ik beloof dat er meer tijd voor onthaasten gaat komen! Mijn allerliefste **Lieke**. Je bent nu iets meer dan een jaar in ons leven. Met jou in de buik heb ik mijn proefschrift afgerond. En nu je de eerste stappen begint te zetten, heb ik nieuwe energie om het dan ook eindelijk te gaan verdedigen. Je bent zo vrolijk, guitig, vol verwondering, open, oprecht en een tikkeltje eigenwijs. Eigenschappen, die je zeker moet bewaren voor later. Dank je wel voor je prachtige lach, je vertrouwen in Daan en mij, en het feit dat je er bent. Ik weet zeker dat succes het resultaat zal zijn van willen en doen! Ik hou van jullie!



Chapter 12

Curriculum Vitae

Harma Johanna Alma was born on March 31st, 1984, in Coevorden, the Netherlands. After graduating from secondary school at the Christelijk Lyceum Veenendaal in 2002, she entered further study in International Hospitality Management at the Stenden University Leeuwarden. In 2005, Harma studied accounting for 6 months at the BI Norwegian School of Management in Oslo, Norway; and in 2006, attained a Bachelor of Business Administration degree. Harma then continued her academic career in Sweden, and in 2008, completed a Master of Science degree in Tourism and Hotel Management at the Graduate Business School, University of Gothenburg. During that period, she was the treasurer on the board of the Graduate School Student Organisation and head of the Social and Sports Committee. For her final thesis, she conducted qualitative research into the reasons why Swedish hoteliers prefer to remain independent from integration into larger hotel chains. Harma then remained in Sweden and continued her career as a housekeeping shift leader at the four-star 532-room Clarion Hotel Stockholm. However, she changed jobs later in 2008 to gain experience as a project manager in business-to-business and consumer marketing at the Netherlands Board of Tourism and Conventions in Stockholm. In that position, she was responsible for the development and execution of various marketing campaigns aimed at Scandinavian tourists, business travellers, travel trade and planners of corporate meetings and congresses. While employed in this role, Harma decided to pursue a lifelong desire to become a medical doctor.

In 2011, Harma started a course in medicine at the University of Groningen, soon graduating in 2013 with a Bachelor's degree after completing a fast-track training programme (so-called *zij-instroom)*. In the period from 2014 to 2018, Harma combined various junior and senior medicine clerkships while conducting research at the department of General Practice and Elderly Care Medicine of the University Medical Center Groningen. This was completed with the Groningen research institute for asthma and COPD in primary care (*GRIAC-PC group*). The present research was also funded by a scholarship of the Junior Scientific Masterclass of the University of Groningen, which enabled Harma to complete the research presented in this thesis. The research was executed in cooperation with Konrad Schultz, medical director of the Klinik Bad Reichenhall, Germany, and under the supervision of Professor Thys van der Molen, Professor Robbert Sanderman and Doctor Corina de Jong at the University Medical Center Groningen. In addition, Harma worked as a student assistant on other research projects and as a student tutor responsible for assisting medical students.

After graduating in medicine in 2018, Harma worked as a physician in the Groningen area of the Netherlands, primarily offering medical services to disabled and elderly patients. In 2020 she continued her medical career as a general practitioner trainee, combining this role with postdoctoral research activities on the topic of anxiety and depression in

children and adolescents in primary care. Harma has significant expertise with quality of life instruments and patient-reported outcome tools as well as research experience in asthma and COPD in primary care settings.

Publications

Alma, H, de Jong, C, Kocks, J and van der Molen, T. Commentary: "Healthcare Professionals' Preferred Efficacy Endpoints and Minimal Clinically Important Differences in the Assessment of New Medicines for Chronic Obstructive Pulmonary Disease" by Dankers M *et al*. in Frontiers in Pharmacology 2020; 10: 1519. *Frontiers in Pharmacology* 2020; 11: 827.

Alma, H, de Jong, C, Jelusic, D *et al.* Baseline health status and setting impacted minimal clinically important differences in COPD: an exploratory study. *Journal of Clinical Epidemiology* 2019; 116: 49-61.

Alma, HJ, de Jong, C, Jelusic, D *et al.* Assessing health status over time: impact of recall period and anchor question on the minimal clinically important difference of COPD health status tools. *Health and Quality of Life Outcomes* 2018; 16(1): 130.

Alma, H, de Jong, C, Tsiligianni, I *et al*. Clinically relevant differences in COPD health status: Systematic review and triangulation. *European Respiratory Journal* 2018; 52(3): 1800412.

Alma, HJ, de Jong, C, Jelusic, D *et al*. Thresholds for clinically important deterioration versus improvement in COPD health status: results from a randomised controlled trial in pulmonary rehabilitation and an observational study during routine clinical practice. *BMJ Open* 2018; 9(6): e025776.

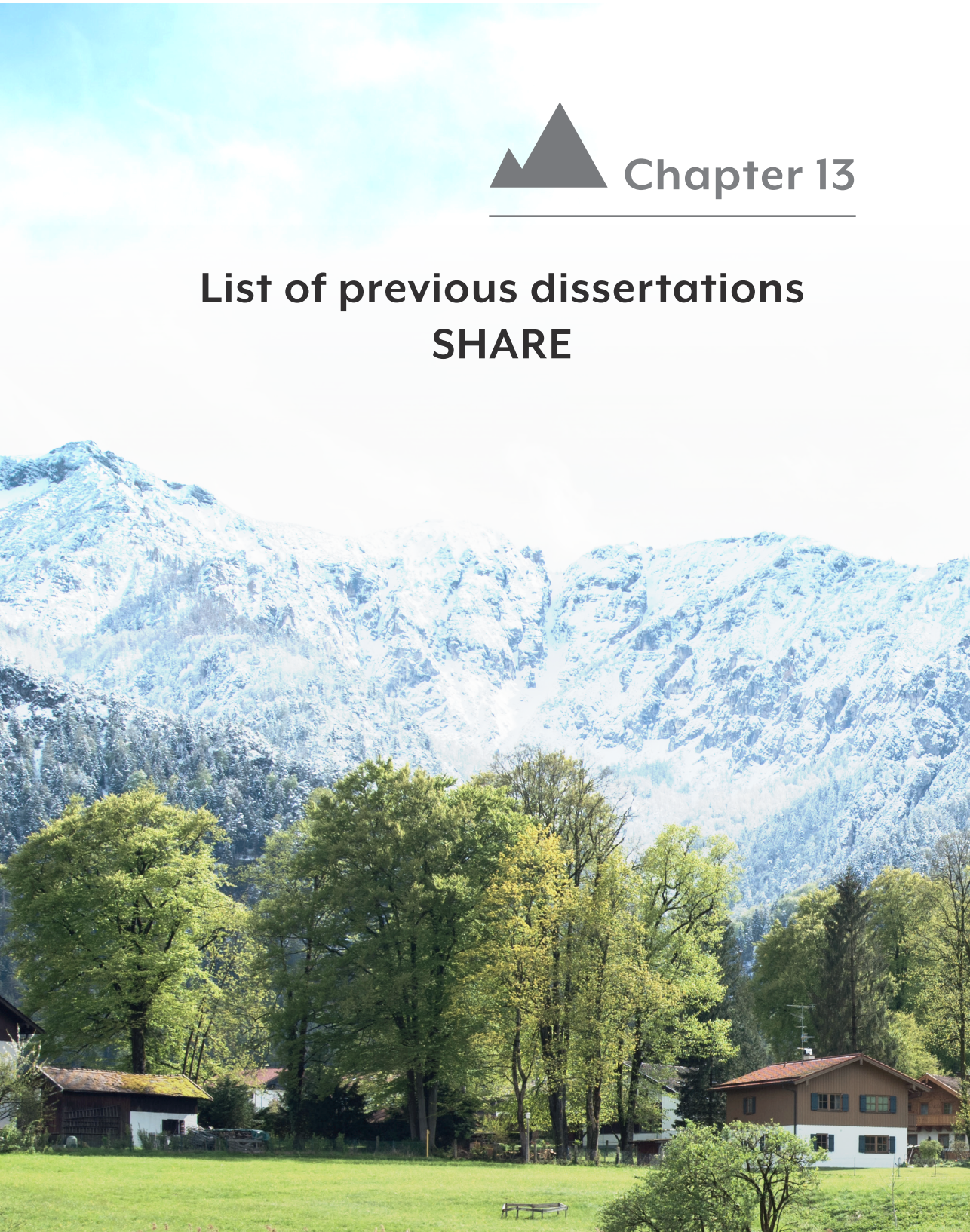
Alma, H, de Jong, C, Jelusic, D *et al.*. Health status instruments for patients with COPD in pulmonary rehabilitation: Defining a minimal clinically important difference. *npj Primary Care Respiratory Medicine* 2016; 26: 16041.

Alma, H. Klinisch relevant en statistisch significant. *Huisarts en Wetenschap* 2015; 58(11): 577.

Blanker, MH, Alma, HJ, Devji, TS *et al.* Determining the minimal important differences in the International Prostate Symptom Score and Overactive Bladder Questionnaire: results from an observational cohort study in Dutch primary care. *BMJ Open* 2019; 9(12): e032795.

Schultz, K, Jelusic, D, Wittmann, M *et al.* Inspiratory muscle training does not improve clinical outcomes in 3-week COPD rehabilitation: Results from a randomised controlled trial. *European Respiratory Journal* 2018; 51(1): 1702000.

Tsiligianni, IG, Alma, HJ, de Jong, C *et al.* Investigating sensitivity, specificity, and area under the curve of the Clinical COPD Questionnaire, COPD Assessment Test, and Modified Medical Research Council scale according to GOLD using St George's Respiratory Questionnaire cutoff 25 (and 20) as reference. *International Journal of Chronic Obstructive Pulmonary Disease* 2016; 11(1): 1045-1052.



Chapter 13

List of previous dissertations SHARE

This thesis is published within the Research Institute SHARE (Science in Healthy Ageing and healthcaRE) of the University Medical Center Groningen / University of Groningen.

Further information regarding the institute and its research can be obtained from our internet site: http://www.share.umcg.nl/

More recent theses can be found in the list below.

(supervisors are between brackets)

2020

Rietkerk W
Tailoring care for older adults; understanding older adults goals and preferences
(prof SU Zuidema, prof JPJ Slaets, prof DL Gerritsen)

Aalst RTA van
An economic assessment of high-dose influenza vaccine
(prof MJ Postma, prof JC Wilschut)

Kuiper D
Implementation of the self-management of well-being interventions; determinants and effects
(prof R Sanderman, prof SA Reijneveld, prof BJM Steverink, dr MM Goedendorp)

Zwertbroek EF
Hypertension in pregnancy; timing of delivery and early screening
(prof CM Bilardo, prof SA Scherjon, dr MTM Faassen, dr H Groen)

Boer P de
Cost-effectiveness of vaccination strategies to protect older adults: focus on herpes zoster and influenza
(prof MJ Postma, prof JC Wilschut)

Buijs-Spanjers KR
Improving delirium education: the role of experiential learning in a serious game
(prof SEJA de Rooij, prof ADC Jaarsma)

Karsten MDA
Women's lifestyle and sexual function; The effects of a preconception intervention in women with obesity
(prof A Hoek, prof TJ Roseboom, dr H Groen)

Koops JC
Understanding nonmarital childbearing. The role of socio-economic background and ethnicity in Europe and North-America.
(prof AC Liefbroer, prof AMH Gauthier)

Wasir R
Moving forward to achieve universal health coverage in Indonesia: progress and challenges
(prof E Buskens, prof MJ Postma, dr W Goetsch)

Kramer T
How to develop a Grand Slam winner...; physical and psychological skills in Dutch junior tennis players
(dr MT Elferink-Gemser, prof C Visscher, dr BCH Huijgen)

Raven D
Where's the need? The use of specialist mental health services in adolescence and young adulthood
(prof AJ Oldehinkel, prof RA Schoevers, dr F Jörg)

Stoter IK
Staying on track; the road to elite performance in 1500 m speed skating
(dr MT Elferink-Gemser, prof C Visscher, prof FJ Hettinga)

Silva Lagos LA
Gestational diabetes mellitus and fetoplacental vasculature alterations; exploring the role of adenosine kinase in endothelial (dys)function
(dr MM Faas, prof T Plösch, prof P de Vos, prof L Sobrevia)

Reints R
On the design and evaluation of adjustable footwear for the prevention of diabetic foot ulcers
(prof ir GJ Verkerke, porf K Postema, dr JM Hijmans)

Bekhuis E
A body-mind map; epidemiological and clinical aspects of the relation between somatic, depressive and anxiety symptomatology
(prof JGM Rosmalen, prof RA Schroevers, dr L Boschloo)

Havinga PJ
Breaking the cycle? Intergenerational transmission of depression/anxiety and opportunaties for intervention
(prof RA Schoevers, prof CA Hartman, dr L Boschloo)

Geer SJ van der
Trismus in head and neck cancer patients
(prof PU Dijkstra, prof JLN Rodenburg, dr H Reintsema)

Salavati N
Preconception environmental factors and placental morphometry in relation to pregnancy outcome
(prof JJHM Erwich, prof RM van der Beek, dr MK Bakker, dr SJ Gordijn)

Fels IMJ van der
Movement, cognition and underlying brain functioning in children
(dr E Hartman, prof C Visscher, prof RJ Bosker, dr J Smith)

Braaksma P
Moving matters for children with developmental coordination
(prof R Dekker, prof CK van der Sluis, dr MM Schoemaker, dr I Stuive)

For earlier theses visit our website